

Instituto Superior de Economia e Gestão
Universidade Técnica de Lisboa

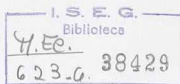
INFLUÊNCIA DAS OBSERVAÇÕES NOS COEFICIENTES ESTIMADOS DO MODELO DE REGRESSÃO LINEAR MÚLTIPLA

Uma Abordagem Fundamentada na Função de Influência

por

JOSÉ Manuel de
Matos **PASSOS**

Dissertação apresentada como requisito
parcial para a obtenção de grau de
Mestre em Métodos Matemáticos
para Economia e Gestão



QA278.2
P37
1992



Instituto Superior de Economia e Gestão

Universidade Técnica de Lisboa

RESERVADO

**INFLUÊNCIA DAS OBSERVAÇÕES NOS
COEFICIENTES ESTIMADOS DO MODELO
DE REGRESSÃO LINEAR MÚLTIPLA**

Uma Abordagem Fundamentada na Função de Influência

José Manuel de Matos Passos

Dissertação apresentada como requisito parcial para a obtenção de grau de
Mestre em Métodos Matemáticos para Economia e Gestão.

Abril de 1992



Ao Professor Doutor Bento Martins, presidente desta
comissão, com o profundo e sincero respeito, pelo
trabalho realizado e a colaboração prestada em
esta ocasião.

Agradeço ainda a todos os colegas e amigos que directa ou
indirectamente contribuíram para a realização deste trabalho.

Quilómetros de estrada e de trabalho, com a ajuda de Deus.

Com os melhores cumprimentos,

A Helena

Ao Ricardo



QUADRO DE MATERIAS

0 INTRODUÇÃO

1 RESUMOS E MATRIZ DE PROJECCÃO, PRIMEIRA

Ao Professor Doutor Bento Murteira, orientador desta dissertação, cabe aqui um profundo e sincero agradecimento pelo auxílio prestado nos diversos momentos da sua realização.

Agradeço ainda a todos os colegas e amigos que directa ou indirectamente contribuíram para a realização deste trabalho.

1.1 Principais Aplicações da Teoria da

Quaisquer erros ou omissões são da inteira responsabilidade

do autor.

1.1.1 Introdução Geral da Teoria da

1.1.2 Introdução Geral da Teoria da

1.1.3 Matriz de Projeccão e Introdução

1.1.3.1 Introdução Geral da Teoria da

1.1.3.2 Introdução Geral da Teoria da

1.1.3.3 Introdução Geral da Teoria da

QUADRO DE MATERIAS E INFLUENCIA DAS

RESERVAÇÕES.

2.1 Noção da Função de Influência.

0 INTRODUÇÃO. 1

1 RESIDUOS E MATRIZ DE PROJECCÃO. PRIMEIRA ABORDAGEM A INFLUÊNCIA. 6

1.1 Modelo e Hipóteses. 6

1.2 Matriz de Projecção. Propriedades. 10

A) Significado da Matriz de Projecção no MRLM. 10

B) Propriedades da Matriz de Projecção. 12

1.3 Primeira Abordagem à Influência. 22

1.3.1 Resíduos e Influência. 23

A) Studentização Interna dos Resíduos. 25

B) Studentização Externa dos Resíduos. 26

1.3.2 Matriz de Projecção e Influência. 29

A) Elementos Diagonais de P . 29

B) Distância de Mahalanobis. 32

C) Elementos Diagonais de P_z . 33

2	FUNÇÃO DE INFLUÊNCIA E INFLUÊNCIA DAS OBSERVAÇÕES.	35
2.1	Noção de Função de Influência.	36
	A) Funcionais Estatísticas e Expansões de Von Mises.	36
	B) Função de Influência de uma Estatística.	41
2.2	Função de Influência do Estimador dos Mínimos Quadrados.	42
	A) Função de Influência Empírica Baseada em n Observações.	45
	B) Função de Influência Empírica Baseada em $(n-1)$ Observações.	45
	C) Função de Influência Amostral.	47
	D) Curva de Sensibilidade.	47
2.3	Medidas de Influência de uma Observação.	51
	A) Medida de Cook.	52
	B) Medida de Welsch-Kuh.	58
	C) Medida de Welsch.	60
	D) Medida de Atkinson.	61
2.4	Medidas de Influência de um Subconjunto de Observações.	63
	A) Abordagem de Cook e Weisberg.	71
	B) Abordagem de Gray e Ling.	72

3	ESTUDO DA INFLUÊNCIA NUM MODELO PARA A PROCURA DE MOEDA EM PORTUGAL.	74
3.1	Algumas Considerações Teóricas sobre a dedução do Modelo para a Procura de Moeda.	74
3.2	Resultados da Estimação do Modelo para a Procura de Moeda.	78
3.3	Identificação das Observações influentes.	84
	A) Influência Individual.	84
	B) Influência Conjunta.	94
4	CONSIDERAÇÕES FINAIS.	106
5	ANEXOS.	108
6	BIBLIOGRAFIA.	114

NOTAÇÃO MAIS FREQUENTE:

- p_{ij} - Elementos da matriz de Projecção, $P = X(X^T X)^{-1} X^T$.
- p_{cij} - Elementos da matriz de Projecção, $P_c = X_c(X_c^T X_c)^{-1} X_c^T$, construída a partir da matriz X em variáveis centradas.
- p_{zij} - Elementos da matriz de Projecção, $P_z = Z(Z^T Z)^{-1} Z^T$, com $Z = [Y \ X]$.
- $p_{(i)ri}$ - Elementos da matriz de projecção, $P_{(i)} = X_{(i)}(X_{(i)}^T X_{(i)})^{-1} X_{(i)}^T$, construída a partir da matriz X sem a observação i .
- $\hat{\sigma}_{(i)}^2$ - Variância dos resíduos calculada sem a observação i .
- r_i - i -ésimo resíduo studentizado (externo).
- t_i - i -ésimo resíduo studentizado (interno).
- $\hat{u}_{(i)i}$ - i -ésimo resíduo preditivo.
- $X_{(i)}$ - Matriz X sem a linha i .
- \bar{x} - Média das colunas de X .
- $\bar{x}_{c(i)}$ - Média das colunas de $X_{(i)}$ em variáveis centradas, $X_{c(i)}$.
- x_i - Linha i da matriz X .
- x_{ci} - Linha i da matriz X_c .
- z_i - Linha i da matriz $Z=[Y \ X]$.



O. INTRODUÇÃO

Quando se está a trabalhar no âmbito do Modelo de Regressão Linear Múltipla (MRLM, daqui em diante), acontece por vezes que alguma ou algumas das observações não se adequam convenientemente aos resultados estimados, no sentido em que diferentes subconjuntos de observações proporcionam estimativas que estão longe de serem semelhantes. Está-se numa situação em que os parâmetros do modelo são demasiados sensíveis a algumas observações, colocando problemas não só relativos à sua estabilidade, mas também à validade das condições de Gauss-Markov, comprometendo consequentemente, qualquer inferência estatística que lhe possa ser feita.

As principais razões deste problema assentam numa incorrecta especificação do modelo e/ou na existência de erros nas observações. Apesar de os modelos resultarem de um corpo teórico coerente, nem sempre se clarifica convenientemente o tipo de relação funcional existente entre as variáveis, por exemplo entre uma especificação linear ou log-linear. Quando a presença deste tipo de observações puder, por exemplo, ser imputada a um problema de escala, uma especificação log-linear pode reduzir ou mesmo eliminar os seus efeitos. Por outro lado, qualquer modelo, como uma simplificação da realidade, procura traduzir os aspectos mais relevantes daquilo que o investigador conhece ou pensa conhecer. Ora, pode muito bem acontecer que essa simplificação seja demasiado forte, no sentido em que as variáveis consideradas como as mais relevantes

sejam em número insuficiente para a explicação do fenómeno em estudo. Em suma, está-se numa situação em que algumas observações aparecem desarticuladas do modelo em consequência de uma má modelização.

De outro modo, se não houver razões para se duvidar da relevância do modelo com que se está a trabalhar então é porque algumas das observações contêm erros.

A presença de tais observações anda associada quer à variável dependente quer às variáveis independentes. No primeiro caso são geralmente conhecidas na literatura pela designação de '*outliers*' e caracterizam-se por originarem resíduos elevados em valor absoluto. No segundo caso, designadas por '*leverages*', são caracterizadas pelo seu afastamento (significativo) face às restantes observações, no espaço das variáveis independentes¹. Em consequência, quanto maior o afastamento de uma observação do seu 'centro de gravidade', menor a importância das restantes observações na determinação do valor dos coeficientes estimados.

'*Outliers*' e '*leverages*' são os responsáveis pela presença de observações influentes, isto é, de observações que quando removidas levam a uma alteração substancial nos resultados da regressão. No entanto nem todas as observações identificadas como

¹ - Rousseeuw e Leroy (1987), considerando o espaço das variáveis do modelo, atribuem-lhe também a designação de '*outliers* na direcção das variáveis independentes'. Quanto maior o afastamento de uma observação da média das observações, no espaço das variáveis independentes do modelo, maior o seu 'efeito alavanca', no sentido em que o hiperplano ajustado, tende a deslocar-se na direcção dessa observação. No limite, quando esse efeito é máximo, o resíduo associado é nulo. Como tal, uma tradução possível para este termo, pelo menos aquela que melhor reflecte o seu significado, poderá ser a de 'Efeito Alavanca da Observação'.

'outliers' e/ou 'leverages' são influentes, como se verá no capítulo 2. Por exemplo, 'outliers' associados a observações que no espaço das variáveis independentes estejam em torno da sua média, apenas influenciam o termo autónomo do modelo, sem afectar os restantes coeficientes.

Sendo pouco desejável a presença de 'outliers' ou 'leverages', é importante para além da simples identificação, conhecer os seus efeitos nos resultados da regressão. Este estudo é conhecido na literatura por 'influência das observações' que, no caso do MRLM, pode ser medida nos coeficientes de regressão estimados, no desvio padrão, na estatística t de Student, entre outros.

Um dos primeiros trabalhos nesta área é devido a Cook (1977), com a apresentação de uma medida para avaliar o efeito de cada observação nos coeficientes estimados do MRLM. Desde então, têm-se desenvolvido vários tipos de abordagens que, em termos genéricos, podem ser classificadas em clássicas e robustas. Quanto à abordagem clássica, as medidas aí propostas agrupam-se, de acordo com a metodologia utilizada, nas seguintes:

- Medidas fundamentadas na noção de função de influência do estimador [Cook (1977,1979), Cook e Weisberg (1980), Belsley, Kuh e Welsch (1980) e Atkinson (1981)];
- Medidas fundamentadas no volume de elipsóides [Andrews e Pregibon (1978), Belsley, Kuh e Welsch (1980) e Cook e Weisberg (1980)];
- Medidas fundamentadas na função de verosimilhança [Cook

e Weisberg (1982)];

- Medidas fundamentadas na estrutura dos valores e vectores próprios da matriz das variáveis independentes [Belsley, Kuh e Welsch (1980)] e
- Medidas fundamentadas em métodos Bayesianos [Johnson e Geisser (1979, 1980)].

A identificação de observações influentes a partir de procedimentos robustos é essencialmente devida aos trabalhos de Rousseeuw (1984) e Hawkins, Bradu e Kass (1984). O primeiro método assenta na regressão robusta, utilizando para o efeito um estimador que se obtém pela minimização da mediana do quadrado dos resíduos (LMS - *Least Median of Squares*). O segundo método parte da noção de conjunto elementar, seguindo posteriormente um procedimento similar ao anterior.

Em face desta multiplicidade, optou-se neste trabalho pelo tratamento da abordagem clássica, em particular pelo estudo da influência fundamentado no conceito de função de influência do estimador dos mínimos quadrados.

No 1º capítulo apresentam-se alguns conceitos que são fundamentais para uma melhor compreensão do significado das medidas de influência. Após uma ligeira apresentação do MRLM, a atenção vai centrar-se no estudo da matriz de projecção, mediante a apresentação de algumas propriedades, bem como dos resíduos do modelo, no sentido de formular os conceitos de '*leverage*' e '*outlier*'², estudando simultaneamente algumas medidas para a sua detecção.

2 - Para um estudo mais aprofundado veja-se Hawkins (1980).

O 2º capítulo inicia-se com uma apresentação breve da noção de função de influência de um estatística, sendo derivada na secção 2.2 a função de influência do estimador dos mínimos quadrados. Esta função é em geral utilizada para estudar a robustez dos estimadores ou, dito de outra forma, a sua sensibilidade à presença de observações que se afastam do seu 'centro de gravidade'. Funções de influência não limitadas correspondem a estimadores que não são robustos, como é o caso dos mínimos quadrados. Atendendo a que esta função é deduzida em termos da população, propõem-se também algumas versões amostrais. Na secção seguinte procura-se fazer a ligação entre o conceito de função de influência amostral e as diferentes medidas de influência das observações nos coeficientes do modelo, com a sua apresentação e discussão.

A finalizar é feito no 3º capítulo uma aplicação a um modelo econométrico para a procura de moeda em Portugal, com o principal objectivo de elucidar e discutir alguns dos métodos apresentados.

1. RESÍDUOS E MATRIZ DE PROJEÇÃO.

PRIMEIRA ABORDAGEM A INFLUÊNCIA.

O principal objectivo deste capítulo é evidenciar o papel desempenhado pelos resíduos e pela matriz de projecção no estudo da influência, bem como a relação destes com outros conceitos, em especial com os de 'outlier' e 'leverage'.

Valores elevados para os resíduos, geralmente associados à presença de 'outliers' na variável dependente do modelo, podem ser ou não influentes em algum ou alguns dos resultados da regressão. De modo análogo, os 'leverages' presentes, associados à existência de valores elevados da diagonal principal da matriz de projecção, podem ser ou não influentes. No capítulo seguinte verifica-se que as medidas de influência se baseiam numa combinação destes efeitos.

1.1. MODELO E HIPÓTESES.

O estudo da influência que aqui se inicia, centra-se no MRLM,

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + u_i, \quad i = 1, 2, \dots, n,$$

ou, em termos matriciais,

$$(1.1) \quad Y = X\beta + U,$$

onde,

$$Y = [y_1 y_2 \dots y_n]^T,$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix},$$

$$\beta = [\beta_1 \beta_2 \dots \beta_k]^T,$$

$$U = [u_1 u_2 \dots u_n]^T.$$

Estas matrizes tem o seguinte significado:

- Y é a matriz coluna das n observações da variável endógena e y_i a i -ésima observação, $i=1,2,\dots,n$;
- X é a matriz das n observações relativas a cada uma das k variáveis exógenas e x_{ij} a i -ésima observação da j -ésima variável exógena, $i=1,2,\dots,n$, $j=1,2,\dots,k$. Para não sobrecarregar a notação vai admitir-se que x_i representa a i -ésima linha de X ;
- β é o vector dos k parâmetros desconhecidos;
- U é o vector das n componentes da variável residual, não observável, e u_i a variável residual correspondente à i -ésima observação, $i=1,2,\dots,n$.

Além da linearidade admitem-se subjacentes ao modelo (1.1)

as seguintes hipóteses,

- independência linear das colunas da matriz X , $r(X)=k^1$;
- a matriz X é não estocástica, observável sem erros²;
- o vector, U , das variáveis residuais, é ortogonal aos regressores no sentido de garantir a consistência do estimador;
- o vector U é aleatório³ com uma distribuição que se supõe conhecida. Em geral, admite-se que $U \sim N(0, \sigma^2 I)^4$.

O estimador clássico dos mínimos quadrados de β obtém-se pela minimização de $(Y - X\beta)^T(Y - X\beta)$ e é dado por,

$$(1.2) \quad \hat{\beta} = (X^T X)^{-1} X^T Y .$$

Verificando-se as hipóteses subjacentes ao modelo, vai apresentar-se, sem demonstração, um conjunto de propriedades que serão utilizadas no decorrer deste capítulo [veja-se Johnston (1984), cap.5]:

1 - O símbolo $r(.)$ representa a característica da matriz que figura dentro de parêntesis.

2 - Esta hipótese será levantada mais adiante para a dedução da funcional do estimador dos mínimos quadrados.

3 - A aleatoriedade da variável residual decorre do seguinte: a presença do vector U no modelo, deve-se simplesmente ao facto de os modelos constituírem uma simplificação da realidade. Como tal, contemplam apenas as variáveis mais relevantes para a explicação do fenómeno em estudo, deixando de fora uma infinidade de outras variáveis com um reduzido poder de explicação e sumariadas em U . É então razoável admitir que estas sejam aleatórias.

4 - A partir da nota anterior e admitindo-se que o número de variáveis insignificantes, excluídas do modelo, e sumarizadas por u_i , são independentes e identicamente distribuídas, então pelo Teorema do Limite Central prova-se que a sua soma tem distribuição que tende para uma normal [veja-se Murteira, Bento (1990), vol.I, pag.386].

1. Para o estimador $\hat{\beta}$ verificam-se as seguintes propriedades,

- a) $\hat{\beta}$ é estimador centrado de β , isto é, $E(\hat{\beta}) = \beta$;
- b) entre os estimadores centrados de β , $\hat{\beta}$, tem variância mínima com, $Var(\hat{\beta}) = \sigma^2(X^T X)^{-1}$;
- c) $\hat{\beta} \sim N^k(\beta, \sigma^2(X^T X)^{-1})$.

2. Os valores estimados da variável endógena obtêm-se por,

$$(1.3) \quad \begin{aligned} \hat{Y} &= X\hat{\beta} \text{ ,} \\ &= PY \text{ ,} \end{aligned}$$

em que $P = X(X^T X)^{-1}X^T$ é a matriz de projecção⁵, e possuem as seguintes propriedades,

- a) $E(\hat{Y}) = E(Y) = X\beta$;
- b) $Var(\hat{Y}) = \sigma^2 P$;
- c) $Cor(\hat{y}_i, \hat{y}_j) = \frac{p_{ij}}{p_{ii}^{1/2} p_{jj}^{1/2}}$;⁶
- d) $\hat{Y} \sim N^n(X\beta, \sigma^2 P)$.

3. Para o vector dos resíduos estimados tem-se,

$$(1.4) \quad \hat{U} = Y - \hat{Y} = (I - P)Y \text{ .}$$

a verificar,

5 - Esta matriz desempenha um papel fulcral no estudo da influência, sendo algumas das suas propriedades estudadas no subcapítulo seguinte.

6 - $Cor(\dots)$ é o coeficiente de correlação das variáveis que figuram dentro de parêntesis [veja-se Murteira (1990), vol.1, pág.198].

- a) $E(\hat{U}) = 0$;
- b) $Var(\hat{U}) = \sigma^2(I - P)$;
- c) $Cor(\hat{u}_i, \hat{u}_j) = \frac{-p_{ij}}{(1-p_{ii})^{1/2}(1-p_{jj})^{1/2}}$;
- d) $\hat{U} \sim N^n(0, \sigma^2(I - P))$.

1.2. MATRIZ DE PROECÇÃO. PROPRIEDADES.

Nesta secção enumeram-se algumas das principais propriedades respeitantes à matriz de projecção, $P = X(X^T X)^{-1} X^T$, dada a sua importância para as diferentes medidas de influência. Uma forma de melhor se compreender o papel da matriz de projecção, P , é mediante a apresentação algébrica do MRLM, que abreviadamente se passa a descrever.

A) Significado da matriz de projecção no MRLM.

Seja, $Y = X\beta$, um sistema linear de n equações a k incógnitas em que o vector Y e as colunas da matriz X são elementos (pontos) de um espaço euclidiano \mathbb{R}^n . Admita-se ainda que $C(X)$ é um subespaço k -dimensional gerado pelas colunas da matriz X . Como em geral Y não pertence à variedade linear $C(X)$ ⁷, aquele sistema de equações é impossível. Nestas condições, o método clássico dos mínimos quadrados propõe uma solução aproximada para β , obedecendo ao critério da

7 - veja-se Pollock (1979), pags 83-84.

minimização do quadrado dos erros. Em termos algébricos esta solução pode ser vista em duas fases: numa primeira, projecta-se ortogonalmente o vector $Y \in \mathbb{R}^n$ no subespaço $C(X)$, $PY = \hat{Y}$, mediante a utilização da matriz de projecção P ; numa segunda, calcula-se o vector $\hat{\beta}$, a partir do sistema consistente, $\hat{Y} = X\hat{\beta}$.

De modo análogo, como P projecta ortogonalmente qualquer vector de \mathbb{R}^n em $C(X)$, conclui-se facilmente que P projecta X nela própria, $PX = X$, e portanto os resíduos estimados do modelo (1.1), \hat{U} , obtidos pela diferença, $Y - \hat{Y}$, podem ser vistos como a projecção ortogonal de Y no subespaço complementar, ortogonal a $C(X)$, por intermédio da matriz de projecção $(I - P)$.

Em suma, o vector $Y \in \mathbb{R}^n$ pode ser decomposto em duas componentes: uma obtida pela projecção ortogonal em $C(X)$, \hat{Y} ; outra, pela projecção ortogonal no subespaço complementar, \hat{U} .

A abordagem ao MRLM pela via algébrica, permite não só realçar o papel da matriz P , como uma matriz de projecção, mas também a sua importância para a qualidade dos resultados da regressão em geral, na medida em que, a relação entre os verdadeiros valores, Y , U e β , e os correspondentes valores estimados, \hat{Y} , \hat{U} e $\hat{\beta}$, dependem exclusivamente desta matriz. Neste sentido, as hipóteses inicialmente formuladas para o vector U só podem aceitar-se para o correspondente vector estimado \hat{U} e consequentemente para \hat{Y} e $\hat{\beta}$, desde que os elementos da matriz P obedeam a um certo tipo de comportamento. A violação destas hipóteses anda geralmente associada à presença de observações extremas em X isto é, a potenciais observações

influentes. Este e outros problemas que fazem depender a performance do modelo relativamente à matriz P , serão abordados no decorrer deste capítulo.

B) Propriedades da matriz de projecção.

Para um estudo mais aprofundado destas propriedades, remete-se o leitor para os trabalhos de Pollock (1979), Belsley, Kuh e Welsch (1980), Huber (1973, 1981) e Chatterjee e Hadi (1988).

Uma primeira propriedade resulta da própria definição de matriz de projecção, dispensando quaisquer comentários adicionais.

Propriedade 1. As matrizes P e $(I - P)$ são simétricas e idempotentes.

Demonstração. Tendo presente a expressão da matriz P , resulta imediatamente que $P^T = P$ e $P^2 = P$. #

Como consequência da propriedade anterior é possível expressar os seus elementos diagonais por,

$$\begin{aligned} (1.5) \quad p_{ii} &= (PP)_{ii} = \sum_{j=1}^n p_{ij}^2, \\ &= p_{ii}^2 + \sum_{j \neq i} p_{ij}^2, \end{aligned}$$

que constitui uma relação importante para a dedução de alguns resultados apresentados mais adiante.

Propriedade 2. Se X é a matriz dos regressores do modelo (1.1), então

$$(a) \quad \text{tr}(P)=k \quad ;$$

$$(b) \quad \text{tr}(I-P)=n-k \quad ;$$

$$(c) \quad \sum_{i=1}^n \sum_{j=1}^n p_{ij}^2 = k \quad .$$

Demonstração. Pelas propriedades do traço de uma matriz [veja-se Luis e Ribeiro (1985), pag.75] tem-se para (a),

$$\begin{aligned} \text{tr}(P) &= \text{tr}[X(X^T X)^{-1} X^T] \quad , \\ &= \text{tr}[(X^T X)^{-1} X^T X] \quad , \\ &= \text{tr}(I) = k \quad , \end{aligned}$$

em que I é a matriz identidade de ordem k . De modo análogo prova-se

(b). Para demonstrar (c) faz-se uso da igualdade (1.5). #

A propriedade seguinte, mostra que a matriz de projecção pode ser decomposta na soma de duas matrizes (de projecção).

Propriedade 3. Seja $X = [X_1 \ X_2]$ onde X_1 é de ordem $n \times q$ com $r(X_1)=q$ e X_2 é de ordem $n \times (k-q)$ com $r(X_2)=k-q$. Considere-se ainda que $P_1 = X_1(X_1^T X_1)^{-1} X_1^T$, $W = (I - P_1)X_2$ e $P_2 = W(W^T W)^{-1} W^T$ são, respectivamente, a matriz de projecção para X_1 , a projecção de X_2 no complemento ortogonal de X_1 e a matriz de projecção para W . Então a matriz de projecção, P , pode ser expressa por $P = P_1 + P_2$.

Demonstração. Veja-se Chatterjee e Hadi (1988), pág 16 e Pollok (1979). #

Um caso particular decorre quando o modelo (1.1) tem termo autônomo. Considerando a partição $X=[e \ X_2]$, em que e é um vector unitário de dimensão n , tem-se

$$P_1 = e(e^T e)^{-1} e^T = \frac{1}{n} e e^T ,$$

$$W = \left(I - \frac{1}{n} e e^T \right) X_2 .$$

Como a matriz W mais não é do que a matriz X em variáveis centradas, X_c , faz-se por conveniência de notação, $W=X_c=CX$ em que

$$(1.6) \quad C = \left(I - \frac{1}{n} e e^T \right) ,$$

é a matriz de centragem. Nestas condições, tem-se pela propriedade 3

$$(1.7) \quad P = \frac{1}{n} e e^T + X_c (X_c^T X_c)^{-1} X_c^T ,$$

ou para o i -ésimo elemento diagonal,

$$(1.8) \quad \begin{aligned} p_{ii} &= \frac{1}{n} + x_{ci} (X_c^T X_c)^{-1} x_{ci}^T , \\ &= \frac{1}{n} + p_{cii} , \end{aligned}$$

com $p_{cii} = x_{ci} (X_c^T X_c)^{-1} x_{ci}^T$ e $x_{ci} = (x_i - \bar{x})$ a linha i da matriz X_c , onde \bar{x} é a média das colunas de X .

As duas propriedades seguintes proporcionam um melhor conhecimento da estrutura da matriz P e sua relação com os resíduos estimados.

Propriedade 4. Os elementos da matriz de projecção são limitados, verificando-se o seguinte, para $i=1,2,\dots,n$, $j=1,2,\dots,n$,

- (a) $0 \leq p_{ii} \leq 1$;
- (b) $-0.5 \leq p_{ij} \leq 0.5$, $i \neq j$;
- (c) se $p_{ii} = 1$ então $p_{ij} = 0$, para todo o $j \neq i$;
- (d) se $p_{ii} = 0$ então $p_{ij} = 0$, para todo o j ;
- (e) se a matriz X contém uma coluna constante, isto é, se o modelo (1.1) tem termo autónomo, tem-se

$$- n^{-1} \leq p_{ii} \leq 1 ,$$

$$- \sum_{j=1}^n p_{ij} = 1 .$$

Demonstração. A demonstração das alíneas (a), (b) e (e) pode ser vista em Chatterjee e Hadi (1988), págs 17-18. Para (c) vai partir-se da igualdade (1.5). Quando $p_{ii} = 1$, tem-se $p_{ii} = p_{ii}^2$ e, pela alínea (a) desta propriedade, a igualdade (1.5) só é verificada para $\sum_{(j \neq i)} p_{ij}^2 = 0$ isto é, quando $p_{ij} = 0$ para todo o j diferente de i . Para demonstrar (d) procede-se de modo análogo. #

Propriedade 5. A relação entre os elementos de P e os resíduos, para $i=1,2,\dots,n$ e $j=1,2,\dots,n$, é dada por:

$$(a) \quad (1 - p_u)(1 - p_{ij}) - p_{ij}^2 \geq 0 \quad ;$$

$$(b) \quad p_u p_{ij} - p_{ij}^2 \geq 0 \quad ;$$

$$(c) \quad p_u + \hat{u}_i^2 / \hat{U}^T \hat{U} \leq 1 \quad .$$

Demonstração. Tendo presente (1.3) c) e (1.4) c), tem-se de imediato (a) e (b), respectivamente. Para uma demonstração de (c), veja-se Chatterjee e Hadi (1988), pag. 19-20. #

As alíneas (c) e (d) da propriedade 4 apresentam duas situações extremas para p_u . Expressando o i -ésimo elemento de \hat{Y} e \hat{U} , por

$$(1.9) \quad \hat{y}_i = p_u y_i + \sum_{j \neq i} p_{ij} y_j \quad ,$$

e

$$(1.10) \quad \hat{u}_i = (1 - p_u) y_i - \sum_{j \neq i} p_{ij} y_j \quad ,$$

respectivamente, quando $p_u = \partial \hat{y}_i / \partial y_i = 0$, então da propriedade 4, alínea (d), deduz-se que $\hat{y}_i = 0$ isto é, \hat{y}_i não é afectado pelo comportamento de y_i , sendo o correspondente residuo estimado igual a y_i . No extremo oposto, quando $p_u = 1$, tem-se, pela mesma propriedade alínea (c), $\hat{y}_i = y_i$ isto é, \hat{y}_i é determinado exclusivamente a partir da observação y_i e, consequentemente, $\hat{u}_i = 0$. Estas situações são, como se disse, extremas, sendo pouco provável a sua ocorrência num problema concreto. Definem entretanto duas polaridades, entre as quais se deve dar particular atenção aos valores elevados de p_u . Por outro

lado, estabelecem uma importante relação entre os resíduos e os elementos da matriz das variáveis explicativas, X , por intermédio de P , como se deduz explicitamente da propriedade 5(c). Em particular, a um valor elevado (reduzido) de p_u corresponde um valor reduzido (elevado) de \hat{u}_i e portanto, a ideia que frequentemente se tem de considerar as 'boas' observações como aquelas que têm associadas resíduos insignificantes, não é em geral verdadeira. Como se viu na introdução, linhas de X muito afastadas da sua média, forçam o hiperplano ajustado a passar por esse ponto.

Esta relação foi inicialmente estudada por Behnken e Draper (1972), confrontando o comportamento das variâncias de \hat{y}_i e \hat{u}_i com o posicionamento de x_i no espaço \mathbb{R}^k . A sua importância está associada à maior ou menor dificuldade na identificação de 'outliers' no modelo (1.1). Se por hipótese se admite que y_i é um 'outlier', o sucesso da sua identificação, a partir da análise dos resíduos, depende do correspondente valor assumido por p_u . Em particular, um valor elevado dificulta a sua identificação em virtude de \hat{u}_i se aproximar de zero. De modo análogo um valor reduzido de p_u pode, erradamente, levar à aceitação de y_i ser um 'outlier'.

Este problema tem levado ao desenvolvimento de diferentes transformações dos resíduos, no sentido de minorar algumas das suas deficiências, como se verá mais adiante.

A relação, que implicitamente se tem analisado, entre os elementos diagonais da matriz P e o posicionamento das linhas de X

no espaço \mathbb{R}^k , pode ser vista a partir de um resultado geométrico interessante. O conjunto de pontos, $x \in \mathbb{R}^k$ que satisfazem a desigualdade,

$$(1.11) \quad x(X^T X)^{-1} x^T \leq \max p_{ii} ,$$

determinam um elipsóide, contendo o ambiente convexo dos n pontos x_i . Em conclusão, linhas de X associadas com pontos próximos da fronteira do elipsóide, correspondem a valores elevados da diagonal principal de P , isto é, a '*leverages*' em X que, segundo Hoaglin e Welsch (1978), podem ter uma influência considerável nos coeficientes estimados do modelo.

Os factores que determinam o posicionamento destes pontos de \mathbb{R}^k no elipsóide (1.11), dependem não só do afastamento face à sua média, mas também da estrutura dos valores e vectores próprios da matriz X . Para tal admita-se que a matriz X tem uma coluna constante e tome-se o modelo (1.1) em variáveis centradas,

$$(1.12) \quad Y_c = X_c \beta_c + U_c ,$$

onde

$$(1.13) \quad \begin{aligned} Y_c &= CY , \\ X_c &= CX , \\ \beta_c &= [\beta_2 \dots \beta_k]^T , \end{aligned}$$

e C tem o significado dado em (1.6).

Como $X_c^T X_c$ é uma matriz simétrica é possível a decomposição

$$(1.14) \quad X_c^T X_c = Q \Lambda Q^T ,$$

em que

$$\Lambda = \text{diag}\{\lambda_2 \dots \lambda_k\}$$

é a matriz dos valores próprios e

$$Q = [q_{.2} \dots q_{.k}] ,$$

a matriz dos vectores próprios associados. Então, apresentando a inversa de (1.14) por,

$$(X_i^T X_i)^{-1} = \sum_{j=2}^k \frac{1}{\lambda_j} q_{.j} q_{.j}^T ,$$

e substituindo posteriormente em (1.8), com $x_{ci} = (x_i - \bar{x})$, obtém-se

(1.15)

$$p_{ii} = \frac{1}{n} + (x_i - \bar{x}) \sum_{j=2}^k \frac{q_{.j} q_{.j}^T}{\lambda_j} (x_i - \bar{x})^T ,$$

$$= \frac{1}{n} + \sum_{j=2}^k \left[\frac{(x_i - \bar{x}) q_{.j}}{\lambda_j^{1/2}} \right]^2 .$$

Seja θ_{ij} o ângulo que o vector próprio $q_{.j}$ faz com $(x_i - \bar{x})$, o coseno desse ângulo vem dado por,

$$\cos(\theta_{ij}) = \frac{(x_i - \bar{x}) q_{.j}}{[(x_i - \bar{x})(x_i - \bar{x})^T]^{1/2}} ,$$

e portanto,

(1.16)

$$p_{ii} = \frac{1}{n} + (x_i - \bar{x})(x_i - \bar{x})^T \sum_{j=2}^k \frac{\cos^2 \theta_{ij}}{\lambda_j} .$$

Por esta decomposição da matriz P conclui-se que p_{ii} é elevado se x_i estiver muito afastado de \bar{x} , e ainda, se estiver substancialmente na direcção de um vector próprio associado a um valor próprio pequeno.

A propriedade que a seguir se apresenta, relaciona os elementos diagonais da matriz P com a dimensão da amostra, estabelecendo a possibilidade de minorar os efeitos da presença de 'leverages' em X , aumentando a dimensão daquela. Antes da sua apresentação, vai dar-se, sob a forma de lema, um resultado que é importante para a demonstração da propriedade seguinte, conhecido como teorema de Sherman-Morrison-Woodbury [veja-se Henderson e Searle (1981)]:

Lema 1 : Sejam A e D matrizes não singulares de ordem k e m respectivamente e B e C de ordem $(k \times m)$. Nestas condições, prova-se que a inversa de $(A + BDC^T)$ existe e é dada por

$$(A + BDC^T)^{-1} = A^{-1} + A^{-1}B(D^{-1} - C^T A^{-1}B)^{-1}C^T A^{-1}.$$

Propriedade 6. Seja X uma matriz de ordem $n \times k$, com $r(X) = k$. Então, para k fixo, os elementos diagonais de P , p_{ii} , $i = 1, 2, \dots, n$, são não crescentes em n .

Demonstração. Seguindo Cook (1979), seja $X_{(i)}$ a matriz que se obtém de X excluindo a linha i^* . O elemento de ordem (r, l) da matriz de projecção calculado a partir de $X_{(i)}$ e representado por $p_{(i)rl}$ é dado por $p_{(i)rl} = x_r(X_{(i)}^T X_{(i)})^{-1} x_l^T$. Aplicando o lema 1 à expressão $(X_{(i)}^T X_{(i)})^{-1} = (X^T X - x_i^T x_i)^{-1}$, com $A = (X^T X)$, $B = x_i^T$, $D=1$ e $C^T = x_i$, chega-se ao seguinte

$$(1.17) \quad p_{(i)rl} = p_{rl} + \frac{p_{ri} p_{il}}{1 - p_{ii}}, \quad \text{para } r, l \neq i.$$

Para o r -ésimo elemento diagonal, quando $l=r$, (1.17) reduz-se para

$$(1.18) \quad p_{(i)rr} = p_{rr} + \frac{p_{ri}^2}{1 - p_{ii}}.$$

Como $0 \leq p_{ii} \leq 1$ e $p_{ri}^2 / (1 - p_{ii}) > 0$ tem-se imediatamente que $p_{(i)rr} > p_{rr}$ para todo o r isto é, os elementos diagonais da matriz de projecção são não crescentes quando se adiciona a linha (observação) i à matriz $X_{(i)}$ #

Relativamente à demonstração anterior admitiu-se

implicitamente a existência da inversa de $X_{(i)}^T X_{(i)}$. Atendendo a que [Veja-se Belsley, Kuh e Welsch (1980), pag.65]

$$(1.19) \quad |X_{(i)}^T X_{(i)}| = (1 - p_{ii}) |X^T X|,$$

8- O índice (i) refere-se à omissão da observação i , qualquer que seja a variável (vector ou matriz).

uma condição necessária e suficiente para que $r(X_{(0)}^T X_{(0)})=k$ é dada pela não nulidade do determinante (1.19) [veja-se Luis e Ribeiro (1985), pag.121]. Ora, como por hipótese $r(X^T X)=k$, (1.19) será não nulo desde que $p_n \neq 1$.

1.3.1. RESÍDUOS E INFLUÊNCIA

Sendo P e $(I-P)$ matrizes idempotentes, a sua estrutura em termos de valores e vectores próprios é apresentada a seguir.

Propriedade 7. As matrizes P e $(I-P)$ têm valores próprios iguais a 0 e 1. Em particular, P tem $(n-k)$ valores próprios iguais a 0 e k iguais a 1. De modo análogo $(I-P)$ tem $(n-k)$ valores próprios iguais a 1 e k iguais a 0.

Demonstração. Veja-se Luis e Ribeiro (1985), cap.6. #

1.3. PRIMEIRA ABORDAGEM A INFLUÊNCIA.

Deu-se a conhecer na introdução deste trabalho que as diferentes medidas de influência das observações, procuram avaliar o seu impacto nos parâmetros estimados do MRLM, a partir da informação proveniente da matriz de projecção e dos resíduos. A presença deste tipo de observações leva a uma alteração substancial dos coeficientes estimados do modelo e a minoração ou a total eliminação dos seus efeitos, nem sempre é tarefa fácil. Os factores que estão na sua origem, devem-se à presença de 'outliers' e/ou 'leverages' e, esta

identificação pode, por vezes, ajudar a minorar os seus efeitos. Neste sentido, vão apresentar-se, separadamente para \hat{U} e P , algumas das medidas mais comuns na literatura.

1.3.1. RESÍDUOS E INFLUÊNCIA.

A impossibilidade de observar a variável aleatória, U , leva a considerar os resíduos estimados, \hat{U} , como a melhor aproximação, dada a relação existente entre \hat{U} e U . Esta relação, implícita em (1.4), é evidente quando se substitui Y pela sua expressão (1.1), obtendo-se

$$(1.20) \quad \begin{aligned} \hat{U} &= (I - P)(X\beta + U) , \\ &= (I - P)U . \end{aligned}$$

Como tal, é desejável que \hat{U} herde tanto quanto possível as propriedades de U . No entanto, devido a (1.20) e a (1.4) b) e c), isto é, à dependência entre \hat{U} e a matriz P , tal situação nem sempre é verificada. Em particular, uma ou mais observações extremas em X , podem levar à violação das hipóteses referentes à homocedasticidade e não autocorrelação dos resíduos. Pelas mesmas razões, a utilização do critério $\max\{\hat{u}_i\}$, para a detecção de eventuais 'outliers' em Y só será fiável, se os correspondentes valores diagonais de P não se afastarem 'significativamente' da sua média. Dito de outra forma, os valores de \hat{u}_i variam negativamente com os correspondentes p_{ii} , $i=1,2,\dots,n$, pelo que \hat{u}_i reduzido pode corresponder a uma observação y_i que é um 'outlier'.

Em suma, mesmo que os elementos de U sejam independentes com igual variância, (1.20) e (1.4) b) e c) mostram que o mesmo não se passa com os elementos de \hat{U} , a não ser numa situação hipotética em que a matriz P seja diagonal.

Neste sentido, para variações acentuadas de p_{ii} , $i=1,2,\dots,n$, Behnken e Draper (1972) aconselham a que, numa análise de diagnóstico, se trabalhe com uma transformação dos resíduos ordinários, no sentido de ultrapassar alguns dos seus inconvenientes. Para o propósito deste trabalho, vai considerar-se um tipo de transformação que consiste, em termos genéricos, na eliminação da sua dependência do factor de escala⁹. Margolin (1977), designa-os por resíduos studentizados. Mais tarde, Davies (1981) subdivide-os em duas classes: studentização interna e externa [referenciado em Cook e Weisberg (1982)]¹⁰.

A) Studentização interna dos resíduos.¹¹

De acordo com o precedente, estes resíduos obtêm-se dividindo os resíduos ordinários por um estimador da sua escala

9 -Note-se por (1.4.d) que \hat{u}_i tem uma distribuição que é dependente de um factor de escala pois, a sua variância é função de σ^2 e de p_{ii} .

10 -A diferença entre studentização interna e externa é devida a Davies (1981) e baseia-se no seguinte: dividindo uma estatística, V , dependente de um factor de escala, por um estimador de escala, T , obtêm-se uma estatística com uma distribuição que não depende de qualquer factor de escala. Se V e T forem construídos a partir das mesmas observações, serão dependentes; em oposição, serão independentes. Estes, designam-se por studentização interna e externa, respectivamente.

11 -Cook e Weisberg (1982) atribuem-lhe a designação abreviada de resíduos studentizados. Outros autores, designam-os por resíduos standartizados.

[Behnken e Draper (1972), pag.103]. Atendendo a que, por (1.4) b), $\text{Var}(\hat{u}_i) = \sigma^2(1 - p_{ii})$, um estimador 'natural', obtém-se substituindo σ^2 por $\hat{\sigma}^2$. Nestas condições tem-se

$$(1.21) \quad t_i = \frac{\hat{u}_i}{\hat{\sigma}(1 - p_{ii})^{1/2}}, \quad i = 1, 2, \dots, n.$$

Ellenberg (1973), [referenciado em Cook e Weisberg (1982), pág.18] deduz a distribuição conjunta para um subconjunto de m ($< n-k$) resíduos studentizados. Para $m=1$ a distribuição de cada t_i , é apresentada no seguinte teorema.

Teorema 1.1. Se $r(X_{(i)}) = k$ e se verificam as hipóteses inicialmente postuladas para U , então para $i=1, 2, \dots, n$

$$\frac{t_i}{n-k} \sim \text{Be} \left[\frac{1}{2}, \frac{1}{2}(n-k-1) \right],$$

onde Be se refere à distribuição beta.

Demonstração: veja-se Cook e Weisberg (1982), pags.18-19. #

Dada a dificuldade em conhecer os valores da distribuição beta, sem o recurso de meios ou 'packages' informáticos, não é frequente a utilização dos resíduos (1.21) na detecção de 'outliers'. A importância de tais resíduos deve-se, essencialmente, ao facto de integrarem algumas das medidas de influência que serão apresentadas no capítulo seguinte.

B) Studentização externa dos resíduos¹²

Esta operação, requer um outro estimador para σ^2 , de forma a que numerador e denominador de (1.21) sejam independentes. Belsley, Kuk e Welsch (1980), Atkinson (1981), Cook e Weisberg (1980, 1982), entre outros, propõem que se estime o parâmetro σ não considerando a observação i ,

(1.22)

$$\begin{aligned}\hat{\sigma}_{(i)}^2 &= \frac{(n-k)\hat{\sigma}^2 - \frac{\hat{a}_i^2}{1-p_{ii}}}{n-k-1} , \\ &= \hat{\sigma}^2 \left(\frac{n-k-t_i}{n-k-1} \right) .\end{aligned}$$

Nestas condições, em alternativa a (1.21) obtém-se,

(1.23)

$$r_i = \frac{\hat{u}_i}{\hat{\sigma}_{(i)}(1-p_{ii})^{1/2}} ,$$

Este tipo de resíduos têm desempenhado importante papel na detecção de 'outliers', sendo vasta a literatura existente nesta área. A derivação de (1.23) pode ser, em alternativa, precedida pela construção de um modelo (*Mean Shift outlier Model*, na literatura inglesa), em que se suspeita que uma, ou um conjunto de observações, por exemplo a observação i , é um 'outlier'. Nestas condições estima-se o modelo,

(1.24)

$$Y = X\beta + d_i\phi + U^* ,$$

¹² -Na literatura, é frequente a utilização de outras designações para este tipo de resíduos. Belsley, Kuh e Welsch (1980), atribuem a terminologia genérica de resíduos studentizados (apenas), ou abreviadamente RSTUDENT. Atkinson (1981) designa-os por resíduos *Jackknife* (dada a similitude com este estimador).

onde d_i é um vector coluna de dimensão n , com o i -ésimo elemento igual à unidade e os restantes nulos e U^* tem as mesmas propriedades de U . Um valor não nulo para o parâmetro ϕ implica que a observação i é um 'outlier'. A regra de teste será então,

$$H_0: E(Y) = X\beta$$

$$H_1: E(Y) = X\beta + d_i\phi$$

Para testar H_0 versus H_1 constroi-se a seguinte estatística,

$$(1.25) \quad F_i = \frac{\hat{U}^T \hat{U} - \hat{U}^{*T} \hat{U}^*}{\hat{U}^{*T} \hat{U}^* / (n - k - 1)}$$

que, quando H_0 é verdadeira, tem distribuição F com 1 e $(n-k-1)$ graus de liberdade. No entanto, como

$$\hat{U}^{*T} \hat{U}^* = \hat{U}^T \hat{U} - \frac{\hat{u}_i^2}{1 - p_u}$$

a expressão (1.25) é similar a (1.23), isto é $F_i = r_i^2$.

O principal problema da utilização deste tipo de resíduos para a detecção de 'outliers', e portanto para a detecção de potenciais observações influentes, prende-se com o facto de sob H_1 , r_i^2 seguir uma distribuição F não central, com parâmetro de não centralidade $\phi^2(1 - p_u)/\sigma^2$, que varia inversamente com p_u , dificultando a distinção entre a distribuição de r_i^2 sob H_0 e H_1 .

13 - (Cox & Wermuth (1984), mostram que se ϕ tem uma distribuição normal, r_i^2 tem uma distribuição F não central com parâmetro de não centralidade $\phi^2(1 - p_u)/\sigma^2$ em virtude de ser uma função de observações com p_u aleatório.

14 - A este respeito ver-se Brandt e Sörbager (1986), pág. 35.

Refira-se ainda que quando a observação candidata a 'outlier' é desconhecida, uma regra prática consiste em tomar $\max_{1 \leq i \leq n} |r_i|$, rejeitando H_0 sempre que $\max_{1 \leq i \leq n} |r_i| > t_{\alpha/2n}(n-k-1)^{1/3}$. Este procedimento limita a probabilidade do erro de 1ª espécie em $\alpha^{1/4}$.

Uma derivação não menos interessante desta estatística pode ser vista em Atkinson (1981). Seja $\hat{u}_{(i)} = y_i - x_i \beta_{(i)}$, $i=1, 2, \dots, n$, o resíduo calculado em i quando y_i não é utilizado na estimação (vulgarmente conhecido por i -ésimo resíduo predictivo). A estatística t para avaliar a diferença $y_i - x_i \beta_{(i)}$ é designada resíduo *Jackknife* por Atkinson e é dada por,

$$(1.26) \quad r_i^* = \frac{\hat{u}_{(i)}}{\sigma_{(i)} [1 + x_i (X_{(i)}^T X_{(i)})^{-1} x_i^T]^{1/2}},$$

onde $X_{(i)}$ é a matriz que se obtém de X omitindo-se a linha i . De acordo com Weisberg (1980),

$$(1.27) \quad x_i (X_{(i)}^T X_{(i)})^{-1} x_i^T = \frac{p_{ii}}{1 - p_{ii}},$$

e, tendo em atenção que os resíduos ordinários e preditivos se relacionam por,

$$\hat{u}_{(i)} = \frac{\hat{u}_i}{1 - p_{ii}},$$

13 - Cook e Weisberg (1980), sugerem que se tome como regra $\max_{1 \leq i \leq n} |r_i| > t_{p_{ii}/2k}(n-k-1)$ em virtude de ser mais potente em observações com p_{ii} elevado.

14 - A este respeito veja-se Kramer e Sonnberger (1986), pág.85.

[veja-se Cook e Weisberg (1982), pag.33], a igualdade (1.26) coincide com (1.23).

Como vai ver-se no capítulo seguinte, a opção por um destes tipos de resíduos tem estado na origem da apresentação de diferentes medidas da influência. A preferência de alguns autores como Belsley, Kuh e Welsch (1980) e Atkinson (1981), por r_i fundamenta-se no seguinte:

- (a) r_i tem uma distribuição facilmente manuseável;
- (b) r_i pode ser interpretado como a estatística t para o teste de significância do parâmetro ϕ no modelo (1.24);
- (c) r_i reflecte mais acentuadamente valores elevados pois, quando $t_i^2 \rightarrow (n-k)$, $r_i^2 \rightarrow \infty$;
- (d) o estimador de σ , $\hat{\sigma}_{(i)}$, é robusto relativamente à observação i .

O debate sobre a utilização de t_i ou r_i nas diferentes medidas de influência é retomado no capítulo 2.

1.3.2 MATRIZ DE PROJECCAO E INFLUENCIA.

A) Elementos diagonais de P

Atendendo a que uma observação influente pode estar associada à presença de 'leverages' em X , uma possível medida, baseia-se no estudo dos elementos diagonais da matriz P . A questão que desde logo se coloca, é a de saber a partir de que valores de

p_u , $i=1,2,\dots,n$, a correspondente observação pode ser considerada como 'leverage' e, portanto, como potencial influente. A resposta a esta questão não é unânime na literatura devido à inexistência de um procedimento matemático que conduza a um único valor crítico para p_u .

Huber (1973,1981), interpretando $1/p_u$ como o número de observações equivalentes que determinam \hat{y}_i , afirma que devem ser consideradas críticas observações correspondentes a um valor de $p_u > 0.5$, sendo a situação ideal quando $p_u < 0.2$ ¹⁵. Portanto, para que uma observação não seja extrema, \hat{y}_i deve ser determinada por um número razoável de observações equivalentes, y_i , nunca inferior a dois.

Um outro procedimento, mais comum na literatura, consiste em tomar a média dos elementos diagonais de P , que pela propriedade 2, alínea (a), é igual a k/n , comparando-os posteriormente com cada p_u , $i=1,2,\dots,n$. Seguindo Behnken e Draper (1972), se a um espaçamento homogêneo das linhas de X correspondem elementos diagonais de P em torno da sua média, Hoaglin e Welsch (1978) propõem, como regra prática, que se considerem extremas, linhas de X associados a $p_u > 2k/n$. Uma forma de chegar ao resultado proposto por Hoaglin e Welsch, consiste em contornar a hipótese da não aleatoriedade da matriz X , supondo que as k variáveis explicativas são independentes

15 - Este valor limite é proposto por Huber(1981),pags.157-162, em estreita relação com o estudo da consistência do estimador. Segundo este, só é possível garantir a consistência de qualquer combinação linear de β , se $\max p_u \rightarrow 0$, quando $n \rightarrow \infty$.

com distribuição normal multidimensional de ordem k ¹⁶. Representando por $\bar{x}_{\epsilon(i)}$ o vector da média dos regressores da matriz $X_{\epsilon(i)}$ isto é, da matriz X em variáveis centradas, a que se excluiu a observação (linha) i , pode-se utilizar a estatística Λ de Wilk para dois grupos onde um deles é constituído por um único ponto, $x_{\epsilon i} = (x_i - \bar{x})$, com a seguinte expressão,

$$(1.27) \quad \Lambda(x_{\epsilon i}) = \frac{|X_{\epsilon}^T X_{\epsilon} - (n-1)\bar{x}_{\epsilon(i)}^T \bar{x}_{\epsilon(i)} - \bar{x}_i^T \bar{x}_i|}{|X_{\epsilon}^T X_{\epsilon}|}.$$

Simplificando o numerador de (1.27) [Veja-se Belsley, Kuh e Welsch (1980), pag.67] obtém-se,

$$(1.28) \quad \begin{aligned} \Lambda(x_{\epsilon i}) &= 1 - \frac{n}{n-1} p_{\epsilon i} \quad , \\ &= \frac{n}{n-1} (1 - p_u) \quad . \end{aligned}$$

Nestas condições, prova-se que [veja-se Rao (1973), pag.570],

$$(1.29) \quad \frac{n-k}{k-1} \left[\frac{1 - \Lambda(x_{\epsilon i})}{\Lambda(x_{\epsilon i})} \right] \sim F(k-1, n-k) \quad ,$$

ou, atendendo a (1.28), que

$$\frac{n-k}{k-1} \frac{p_u - 1/n}{1 - p_u} \sim F(k-1, n-k) \quad .$$

16 - Esta hipótese é apenas utilizada como orientação para a determinação de um valor crítico para p_u .

Belsley, Kuh e Welsch (1980), mostram que para $k > 10$ e $n - k > 50$, o valor da estatística F a 95% é inferior a 2 pelo que $2k/n$ constitui um bom valor de referência, a partir do qual uma dada observação, x , é considerada como extrema.

B) Distância de Mahalanobis.

No caso de o modelo (1.1) ter termo autónomo, pode-se medir o afastamento das linhas de X , entendidas como pontos de \mathbb{R}^k , face à média das suas colunas, por intermédio da chamada distância de Mahalanobis. Viu-se anteriormente que um procedimento possível, para medir o afastamento de x_i face a \bar{x} , pode ser dado pelo cálculo de,

$$\frac{1}{(n-1)}(x_i - \bar{x})^T (X_c^T X_c)^{-1} (x_i - \bar{x}) .$$

No entanto, se o objectivo for o de medir o afastamento de x_i não relativamente à média das colunas de X , mas às restantes linhas de X (entenda-se, pontos de \mathbb{R}^k), é natural excluir x_i quer de \bar{x} quer de $(n-1)^{-1}(X_c^T X_c)^{-1}$. Sendo $\bar{x}_{c(i)} = (n-1)^{-1}e^T X_{c(i)}$ a média das colunas de $X_{c(i)}$, isto é, das colunas de X_c sem a linha i , a distância de Mahalanobis é dada por,

$$(1.30) \quad MD_i = (n-2)(x_i - \bar{x}_{c(i)}) \left[X_{c(i)}^T \left(I - \frac{1}{(n-1)} ee^T \right) X_{c(i)} \right]^{-1} (x_i - \bar{x}_{c(i)})^T .$$

Atendendo a que a média das colunas de X_c é nula, $n^{-1}e^T X_c = \bar{x}_c = 0$, a média das colunas de $X_{c(i)}$, $\bar{x}_{c(i)}$, é não nula, com [veja-se Chatterjee e Hadi (1988), pag.102],

$$\bar{x}_{c(i)} = \frac{1}{n-1} e^T X_{c(i)} = -\frac{1}{n-1} x_{ci} \quad ,$$

pelo que, a expressão (1.30) simplifica-se para,

$$(1.31) \quad MD_i = \frac{n(n-2)p_u - 1/n}{n-1 \quad 1-p_u} \quad , \quad i = 1, 2, \dots, n \quad .$$

A relação entre Λ_i e MD é, para o mesmo $i=1, 2, \dots, n$, dada por,

$$(1.32) \quad \frac{1 - \Lambda_i}{\Lambda_i} = \frac{(n-1)}{n(n-2)} MD \quad ,$$

podendo-se utilizar, como referência, a estatística de Wilk para a determinação de um limite superior para MD .

C) Elementos diagonais de P_x .

Uma das principais desvantagens das medidas anteriores, prende-se com o facto de estas desprezarem a informação contida em Y . Esta informação pode ser introduzida, alongando a matriz X com o vector Y isto é, construindo a matriz $Z = [Y \ X]$, de ordem $n \times (k+1)$. Para a matriz Z , a correspondente matriz de projecção, P_z , será

$$(1.33) \quad P_z = Z(Z^T Z)^{-1} Z^T \quad ,$$

e atendendo à propriedade 3, a matriz P_z pode ser decomposta na soma de duas matrizes, obtendo-se

$$(1.34) \quad P_z = P + \frac{(I-P)YY^T(I-P)}{Y^T(I-P)Y},$$

$$= P + \frac{\hat{Q}\hat{Q}^T}{\hat{Q}^T\hat{Q}},$$

ou para o i -ésimo elemento diagonal,

$$(1.35) \quad p_{zi} = p_{ii} + \frac{\hat{u}_i^2}{\hat{Q}^T\hat{Q}}.$$

Pela propriedade 5, os p_{zi} são também inferiores à unidade e serão tanto maiores quanto maiores \hat{u}_i e/ou p_{ii} . Para a determinação de um valor 'crítico', pode-se proceder de forma análoga à que se fez para os elementos diagonais de P .

Esta medida proporciona de alguma forma, uma ideia da influência de cada par z_i nos resultados da regressão. No entanto, os resultados onde essa influência se faz notar não é clara. Por outro lado, a sua dedução carece de fundamentação teórica. Uma forma natural de abordar a influência parte da noção de função de influência teórica que se passa a descrever no capítulo seguinte.

2. FUNÇÃO DE INFLUÊNCIA E INFLUÊNCIA DAS OBSERVAÇÕES

O conceito de função de influência foi apresentado e estudado por Hampel (1968,1974), em estreita relação com a noção de robustez¹. Muito genericamente, esta função é construída pontualmente e procura medir a influência que uma particular observação tem sobre o estimador. Um estimador será classificado de robusto se esta função for limitada, isto é, se as estimativas obtidas não variarem significativamente, quando se é confrontado com a presença de observações extremas.

Uma grande parte deste trabalho vai-se basear na noção de função de influência, não com o objectivo de avaliar ou propor estimadores robustos, mas de estudar as consequências da presença de observações extremas, na performance dos resultados estimados, quando estes se obtêm a partir de estimadores que não são robustos.

Os fundamentos deste conceito, remontam ao trabalho iniciado por Von Mises (1947), com a criação da classe das funcionais estatísticas. Neste sentido inicia-se com uma breve exposição do seu método, não só pela razão apontada mas, essencialmente, por permitir uma melhor compreensão da noção de influência.

¹-Inicialmente, Hampel designa-a por curva de influência. Só mais tarde se faz a sua generalização para o caso n dimensional.

2.1. NOÇÃO DE FUNÇÃO DE INFLUÊNCIA.

A) Funcionais estatísticas e expansões de Von Mises.

Seja, x_1, x_2, \dots, x_n , uma amostra casual de n variáveis aleatórias, com domínio no espaço amostral Ω^2 , independentes e identicamente distribuídas (iid), com distribuição F_0^3 e F_n a correspondente função de distribuição empírica, dada por

$$(2.1) \quad F_n(x) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} ,$$

onde δ_{x_i} é a distribuição que atribui toda a massa de probabilidade ao ponto x_i . Se uma estatística, $T_n = T_n(X_1, X_2, \dots, X_n)$, pode ser expressa como funcional, T , da função de distribuição empírica, $T_n = T(F_n)$, onde T não depende de n , diz-se que T é uma funcional estatística. Relativamente à estatística T_n , vai-se admitir a sua consistência no sentido de Fisher. Esta exigência, de modo algum limitativa para o propósito deste trabalho, tem a vantagem de permitir trabalhar com a mesma funcional, T , quer para o parâmetro a estimar, $T(F) = \theta$, quer para o estimador, $T(F_n) = \hat{\theta}$.

Definindo \mathcal{F} como o conjunto de todas as funções de distribuição (teóricas e empíricas), então a funcional T pode ser interpretada como uma aplicação que a cada elemento de \mathcal{F} faz

2-Para simplificar, admita-se que o espaço amostral coincide ou é subconjunto de \mathcal{R} .

3- Por comodidade de notação vai-se suprimir o índice de F . Daqui por diante, sempre que se refere a uma função de distribuição, admite-se a sua dependência de um parâmetro desconhecido.

corresponder um número real. Com base nestes conceitos, Von Mises procura linearizar localmente a estatística $T(F_n)$, aproximando-a por uma funcional linear, numa vizinhança de F , isto é, considerando uma expansão em fórmula de Taylor, com o objectivo de estudar a sua distribuição assintótica.

Em traços gerais, se \mathcal{F} é um conjunto convexo então, quaisquer que sejam as funções de distribuição F e G de \mathcal{F} , com $T(F)$ e $T(G)$ finitas, a funcional,

$$(2.2) \quad T(F_t) = T[F + t(G - F)] ,$$

é, sob certas condições, uma função contínua de t , podendo expandir-se em fórmula de Taylor de ordem m , baseada no ponto fixo F ($t=0$), isto é,

$$(2.3) \quad T(F_t) = T(F) + tT'_F(G - F) + \frac{t^2}{2}T''_F(G - F) + \dots + R_m(G - F) ,$$

em que $T'_F(G - F)$, $T''_F(G - F), \dots$, são as sucessivas derivadas de Von Mises da funcional $T(F_t)$. Em particular, a primeira derivada de Von Mises é dada por,

$$(2.4) \quad \begin{aligned} T'_F(G - F) &= \frac{d}{dt} T[F + t(G - F)]_{t=0} , \\ &= \lim_{t \rightarrow 0} \frac{T[F + t(G - F)] - T(F)}{t} , \end{aligned}$$

se existir uma função real ϕ_F , independente de G , tal que

$$(2.5) \quad T'_F(G - F) = \int \phi_F(x) d(G - F)(x) ,$$

onde o integral é definido em Ω . A função ϕ_F na equação (2.5) é única, a menos de uma constante aditiva, e é usualmente normalizada de forma a que,

$$(2.6) \quad \int \phi_F(x) dF(x) = 0,$$

obtendo-se,

$$(2.7) \quad T'_F(G-F) = \int \phi_F(x) dG(x).$$

O termo $T[F+t(G-F)]$, designa-se por funcional da função de distribuição F contaminada por G e, de acordo com (2.4), a derivada de Von Mises mais não é do que a derivada de T em F na direcção de G ⁴.

A derivada de Von Mises, definida por (2.4), aparece por vezes referida com a designação de derivada de Gâteaux. Em termos formais são idênticas mas, no entanto, os fundamentos divergem.

Designa \mathcal{M} o conjunto das medidas de probabilidade em \mathcal{X} e defina-se neste conjunto uma topologia fraca⁵, isto é, a mais fraca topologia na qual todas as funcionais lineares do tipo,

$$(2.8) \quad T(F) = \int \phi_F(x) dF(x),$$

4- Von Mises faz uma adaptação da derivada de Volterra às funcionais estatísticas.

5- Uma forma habitual de introduzir a topologia fraca consiste em admitir que \mathcal{M} é um espaço métrico, introduzindo posteriormente uma métrica geradora desta topologia. Huber(1981), faz uma discussão em torno das métricas de Lévy, Prohorov e Lipschitz limitada.

são contínuas, para qualquer função, ϕ limitada e contínua. Então a derivada de Von Mises é contínua em \mathcal{M} . Adicionalmente, sendo \mathcal{M}^* o conjunto de todas as medidas de sinal limitadas (*signed measures*, na literatura inglesa), munido da topologia fraca, se a funcional estatística puder ser estendida a \mathcal{M}^* , então as derivadas de Von Mises e de Gâteaux serão coincidentes neste espaço⁶.

Quanto maior a proximidade entre as distribuições F_t e F , menor a magnitude dos termos de ordem superior de (2.3). Portanto, para G e F 'relativamente próximas', a expansão (2.3), para $t=1$, pode ser truncada nos seus primeiros termos. Para uma aproximação linear, (2.3) reduz-se a,

$$(2.9) \quad T(G) = T(F) + \int \phi_F(x) dG(x) + R(G-F) \quad .$$

Substituindo em (2.9) a função de distribuição G pela função de distribuição empírica, F_n , tem-se,

$$(2.10) \quad T(F_n) = T(F) + \int \phi_F(x) dF_n(x) + R(F_n - F) \quad .$$

Com base nesta expressão Von Mises desenvolve um estudo da distribuição assintótica da funcional $T(F_n)$, baseando-se no facto de o termo linear,

6- Para topologias mais fortes, como é o caso da topologia induzida pela métrica de Hellinger, estas duas derivadas não são comparáveis [veja-se Fernholz (1983), pag.18].

$$T'_F(F_n - F) = \int \phi_F(x) dF_n(x) , \\ = \frac{1}{n} \sum_{i=1}^n \phi_F(x_i) ,$$

poder ser dado por uma soma de variáveis aleatórias iid. No entanto, a validade deste desenvolvimento, bem como a normalidade assintótica da funcional, dependem do comportamento do resto, $R(F_n - F)$. Tal só será possível se

$$(2.11) \quad n^{1/2} R(F_n - F) \xrightarrow{p} 0 ,$$

em que o índice superior 'p' significa convergência em probabilidade. Por sua vez, esta última está directamente relacionada com as propriedades da derivada da funcional. Basicamente, encontram-se na literatura três noções distintas de derivada: derivada de Gâteaux (ou derivada fraca) e de Fréchet e Hadamard (ou derivadas fortes)⁷. Cada uma, prende-se com diferentes noções de convergência do resto para zero, à medida que F_n converge para F . Utilizando a derivada de Gâteaux (ou a de Von Mises) é suficiente para (2.11) provar que,

$$n^{1/2} \sup_{0 \leq t \leq 1} \left| \frac{d^2}{dt^2} T[F + t(F_n - F)] \right| \xrightarrow{p} 0 .$$

O principal inconveniente desta abordagem está na necessidade de a verificação de (2.11) passar pelo cálculo de uma derivada de ordem superior à primeira, que nem sempre se apresenta tarefa fácil. Em

7- Para uma discussão destes conceitos, os trabalhos de Serfling (1980) e Huber (1981), para as derivadas de Gâteaux e Fréchet, e Fernholz (1983), para a de Hadamard, constituem boas referências.

consequência têm-se utilizado outras noções mais fortes de derivada como é o caso da derivada de Fréchet e de Hadamard. Para o propósito deste trabalho, em particular para a definição de função de influência, é suficiente a derivada de Gateaux (Von Mises).

B) Função de influência de uma estatística.

O trabalho iniciado por Von Mises, de que se deu a conhecer alguns dos passos mais relevantes, está na base da noção de função de influência proposta por Hampel.

Se a funcional T é diferenciável à Gateaux, admitindo a validade do desenvolvimento (2.10), isto é, admitindo que $R = O_p(n^{-1/2})$, deduz-se facilmente desta expressão que o erro cometido na estimação de $T(F)$ a partir de $T(F_n)$ é aproximadamente,

$$(2.12) \quad \frac{1}{n} \sum_{i=1}^n \phi_F(x_i) ,$$

pelo que $\phi_F(x_i)$ pode ser interpretado como a contribuição de x_i para o erro aproximado, $T(F_n) - T(F)$ ou a influência de x_i sobre o mesmo erro. Uma forma alternativa de apresentar a noção de função de influência, mais comum na literatura, sem ligação directa ao trabalho de Von Mises, é dada na seguinte definição:

Definição. A função de influência da funcional T em F é a derivada de Gateaux de T em F , na direcção de δ_x , isto é

$$(2.13) \quad IC(x; F, T) = \lim_{t \rightarrow 0} \frac{T[F + t(\delta_x - F)] - T(F)}{t}.$$

Esta função é definida pontualmente e, para n elevado, mede o efeito que a observação adicional x tem sobre a estatística T_n^8 .

2.2. FUNÇÃO DE INFLUÊNCIA DO ESTIMADOR DOS MÍNIMOS QUADRADOS.

A dedução da função de influência do estimador dos mínimos quadrados, β , como uma particularização de (2.13), tem por principal objectivo fundamentar as diversas medidas de influência das observações a apresentar na secção 2.3. Com base no MRLM (1.1) e de acordo com Hinkley (1977), vai admitir-se o caso mais geral em que o vector $z = (x, y)$, de ordem $1 \times (k+1)$, é aleatório, possuindo distribuição conjunta F , com vector de médias nulas e matriz de segundos momentos dadas por,

$$(2.14) \quad E_F(z^T z) = \begin{bmatrix} \Sigma_{xx}(F) & \Sigma_{xy}^T(F) \\ \Sigma_{xy}(F) & \sigma_{yy} \end{bmatrix}.$$

Nestas condições o estimador de β obtém-se de modo similar a (1.2) por,

$$(2.15) \quad \int x^T (y - x\beta) dF(x, y) = 0,$$

⁸-As suas propriedades são discutidas detalhadamente em Hampel (1974).

cujas soluções fornece a funcional β .

A consideração de $E(z)=0$ visa apenas a simplificação dos cálculos não prejudicando os resultados subsequentes. Por outro lado, a aleatoriedade de $z=(x,y)$ é apenas uma generalização e não invalida que numa aplicação concreta se considere os regressores como não estocásticos.

Se $\Sigma_{xx}(F)$ em (2.14) é uma matriz de ordem k não singular, a funcional correspondente ao estimador dos mínimos quadrados representa-se por,

$$(2.16) \quad T(F) = \Sigma_{xx}^{-1}(F) \Sigma_{xy}^T(F) \quad .$$

Seja δ_x a função de distribuição de massa 1 no ponto $z=(x,y)$ e $F_t = (1-t)F + t\delta_x$ a função de distribuição F contaminada por δ_x . Então

$$(2.17) \quad T(F_t) = \Sigma_{xx}^{-1}(F_t) \Sigma_{xy}^T(F_t) \quad .$$

Atendendo à linearidade da funcional, os dois factores do segundo membro da expressão anterior podem ser dados por,

$$(2.18) \quad \begin{aligned} \Sigma_{xx}(F_t) &= \Sigma_{xx}[(1-t)F + t\delta_x] \quad , \\ &= (1-t) \left[\Sigma_{xx}(F) + \frac{t}{1-t} x^T x \right] \quad ; \end{aligned}$$

$$(2.19) \quad \begin{aligned} \Sigma_{xy}^T(F_t) &= \Sigma_{xy}^T[(1-t)F + t\delta_x] \quad , \\ &= (1-t) \Sigma_{xy}^T + t x^T y \quad . \end{aligned}$$

Por (2.18) constata-se facilmente que a passagem de $\Sigma_{xx}(F)$ para $\Sigma_{xx}(F_t)$ é equivalente à adição de um ponto x de massa discreta, com ponderação

$t/(1-t)$. Esta expressão, é dada como uma soma de duas matrizes e portanto a sua inversão será feita a partir do lema 1, apresentado na secção 1.2. Fazendo $A = \Sigma_{xx}(F)$, $B = C = x$ e $D = I$, $\Sigma_{xx}^{-1}(F_t)$ vem dado por,

$$(2.20) \quad \Sigma_{xx}^{-1}(F_t) = \frac{t}{1-t} \left[\Sigma_{xx}^{-1}(F) - \frac{t}{1-t} \Sigma_{xx}^{-1}(F) x^T \left(1 + \frac{t}{1-t} x \Sigma_{xx}^{-1}(F) x^T \right)^{-1} x \Sigma_{xx}^{-1}(F) \right].$$

Substituindo (2.20) e (2.19) em (2.17), e este último conjuntamente com (2.16) na definição de função de influência (2.13), simplificando e passando ao limite obtém-se a função de influência do estimador de β ,

$$(2.21) \quad IC[(x, y), T, F] = \Sigma_{xx}^{-1}(F) x^T (y - x T(F)).$$

Se o objectivo for o de analisar a influência numa combinação linear de $q < k$ elementos do estimador de β , a dedução da função de influência para essa combinação processa-se de modo análogo. Fazendo $\psi = L\beta$, onde L é uma matriz de ordem $(q \times k)$, com $r(L) = q$, a função de influência para $\hat{\psi}$ é dada por

$$(2.22) \quad IC[(x, y), LT, F] = L \cdot IC[(x, y), T, F].$$

A imagem do que foi dito no capítulo da introdução, confirma-se por (2.21) que a função de influência não é limitada, pois depende da diferença $y - xT(F)$. Por outro lado, mesmo que esta diferença seja reduzida, a influência de um particular $z = (x, y)$ pode ser significativa desde que o valor assumido por x se afaste consideravelmente de $E(x)$, em especial, quando este afastamento se dá na direcção de um vector próprio correspondente a um pequeno valor próprio de $\Sigma_{xx}(F)$.

A expressão (2.21) mede a influência no estimador, decorrente da adição de um ponto $z=(x,y)$ de massa discreta, em termos populacionais. Numa aplicação concreta, a amostra disponível é de dimensão finita, ou mesmo reduzida, sendo conveniente a utilização de uma aproximação àquela expressão. As aproximações que vão ser abordadas podem-se subdividir em dois grupos: um primeiro, parte da definição de função de influência (2.13), onde se substitui a função de distribuição teórica, F , pela distribuição da amostra; um segundo grupo é construído de modo análogo, com a diferença de se omitir o limite de (2.13). Para o efeito, e em quaisquer dos casos, (2.13) ou (2.21) devem ser modificadas substituindo (x,y) por (x_i, y_i) , $i=1,2,\dots,n$, e os parâmetros por estatísticas.

A) Função de influência empírica baseada em n observações (EIC_i)

A função de influência empírica baseada em n observações, constitui uma aproximação directa de (2.21) e obtém-se substituindo F pela função de distribuição empírica, F_n , onde $T(F_n)=\beta$. Nestas condições, $\Sigma_{xx}(F_n)=n^{-1}(X^T X)$ e a EIC é dada por

$$(2.23) \quad EIC_i = EIC(x_i, y_i) = n(X^T X)^{-1} x_i^T \hat{u}_i, \quad i=1,2,\dots,n.$$

Esta expressão, mede o efeito de uma perturbação infinitesimal de F_n no ponto x_i e corresponde ao *Jackknife* infinitesimal de Jaeckel (1972), discutido em Miller (1974).

B) Função de influência empírica baseada em $(n-1)$ observações ($EIC_{(i)}$)

A função de influência empírica baseada em $(n-1)$ obser-

vações, procura evidenciar a influência que a observação i tem no estimador, mediante a sua exclusão, e obtém-se de (2.21) tomando $F = F_{n(i)}$, $\Sigma_{xx}(F_{n(i)}) = (n-1)^{-1}(X_{(i)}^T X_{(i)})$ e $T(F_{n(i)}) = \hat{\beta}_{(i)}$ onde $F_{n(i)}$, $X_{(i)}$ e $\hat{\beta}_{(i)}$ são, respectivamente, a função de distribuição empírica, a matriz dos regressores e o estimador dos mínimos quadrados, calculados sem a observação i . Então

$$(2.24) \quad EIC_{(i)}(x_i, y_i) = (n-1)(X_{(i)}^T X_{(i)})^{-1} x_i^T (y_i - x_i \hat{\beta}_{(i)}) \quad , \quad i = 1, 2, \dots, n \quad .$$

O termo $(y_i - x_i \hat{\beta}_{(i)}) = \hat{u}_{(i)i}$ é o i -ésimo resíduo preditivo. Expressando $\hat{\beta}_{(i)}$ por

$$\begin{aligned} \hat{\beta}_{(i)} &= (X_{(i)}^T X_{(i)})^{-1} X_{(i)}^T Y_{(i)} \quad , \\ &= (X^T X - x_i^T x_i)^{-1} (X^T Y - x_i^T y_i) \quad , \end{aligned}$$

ou, utilizando o lema 1, por

$$(2.25) \quad \hat{\beta}_{(i)} = \hat{\beta} - (X^T X)^{-1} x_i^T \frac{\hat{u}_i}{1 - p_{ii}} \quad ,$$

e aplicando posteriormente o mesmo lema ao termo $(X_{(i)}^T X_{(i)})^{-1}$ de (2.24),

a $EIC_{(i)}$ pode ser dada, de uma forma mais conveniente, pela seguinte expressão equivalente,

$$(2.26) \quad EIC_{(i)}(x_i, y_i) = (n-1) \frac{(X^T X)^{-1} x_i^T \hat{u}_i}{(1 - p_{ii})^2} \quad , \quad i = 1, 2, \dots, n \quad .$$

Estas duas últimas aproximações medem a influência em termos infinitesimais e portanto, a sua aplicação está limitada pelo facto de ser necessário a utilização de uma amostra infinitamente

grande para a estimação de F_n e $F_{n(t)}$. De forma a obviar este inconveniente vai-se suprimir o limite de (2.13), cujas aproximações se passam a descrever;

C) Função de influência amostral (SIC_i)

A função de influência amostral deduz-se a partir da definição (2.13), fazendo $(x, y) = (x_i, y_i)$, $F = F_n$, $t = -(n-1)^{-1}$, omitindo-se o limite em t e é dada por

$$(2.27) \quad SIC_i = -(n-1) \left[T \left(\frac{n}{n-1} F_n - \frac{1}{n-1} \delta_{z_i} \right) - T(F_n) \right] ,$$

ou, atendendo a que,

$$T \left(\frac{n}{n-1} F_n - \frac{1}{n-1} \delta_{z_i} \right) = T(F_{n(t)}) ,$$

por,

$$(2.28) \quad SIC_i = (n-1) [T(F_n) - T(F_{n(t)})] , \quad i = 1, 2, \dots, n .$$

Substituindo em (2.28) $T(F_n)$ e $T(F_{n(t)})$ por $\hat{\beta}$ e $\hat{\beta}_{(t)}$, respectivamente, obtém-se a seguinte expressão para a SIC

$$(2.29) \quad SIC_i = (n-1) (\hat{\beta} - \hat{\beta}_{(t)}) , \quad i = 1, 2, \dots, n .$$

D) Curva de sensibilidade (SC_i)

A curva de sensibilidade obtém-se de um modo similar da anterior, fazendo $F = F_{n(t)}$ e $t = n^{-1}$ em (2.13), e é dada por

$$(2.30) \quad \begin{aligned} SC_i &= n \left[T \left(\frac{n-1}{n} F_{n(i)} + \frac{1}{n} b_{zi} \right) - T(F_{n(i)}) \right] , \\ &= [T(F_n) - T(F_{n(i)})] , \quad i = 1, 2, \dots, n . \end{aligned}$$

Como $T(F_n) = \beta$ e $T(F_{n(i)}) = \beta_{(i)}$ tem-se,

$$(2.31) \quad \begin{aligned} SC_i &= n(\beta - \beta_{(i)}) , \\ &= n \frac{(X^T X)^{-1} x_i^T \hat{u}_i}{1 - p_u} , \quad i = 1, 2, \dots, n , \end{aligned}$$

com $\beta_{(i)}$ dado por (2.25). Em traços gerais, a SIC e a SC são idênticas, podendo ser interpretadas como a variação ocorrida nos coeficientes estimados quando se exclui uma observação da amostra.

Comparando estas quatro aproximações à função de influência, constata-se que a principal diferença, à parte da dimensão da amostra, está na potência do denominador de cada uma daquelas expressões. Em consequência,

$$\frac{EIC_i}{n} \leq \frac{SC_i}{n} = \frac{SIC_i}{n-1} \leq \frac{EIC_{(i)}}{n-1} .$$

Como vai ver-se na secção seguinte, de entre as versões finitas da função de influência, a que mais se difundiu neste tipo de problemas foi a SIC_i , não só por permitir uma interpretação mais sugestiva, mas também por corresponder a uma situação intermédia entre a EIC_i e a $EIC_{(i)}$. A interpretação mais sugestiva, deriva do facto de esta poder ser vista como a variação ocorrida no estimador, devida à não consideração de uma observação, que é em geral

a ideia subjacente no espírito do analista, quando procura analisar a influência individual de uma observação. Por outro lado, sendo p_u uma medida da existência de 'leverages', a EIC_i peca por ser pouco sensível a este tipo de observações, não as detectando em muitos problemas concretos. Em oposição, a $EIC_{(i)}$ é demasiado sensível. Portanto, a situação intermédia, mais moderada, tem-se apresentado como a mais comum neste tipo de problemas. Note-se entretanto que não há, do ponto de vista teórico, nada que justifique a opção por uma ou outra aproximação. É por vezes a natureza do problema ou o tipo de análise pretendido que está na base da utilização de uma particular função de influência amostral.

Uma abordagem alternativa para a derivação da EIC_i , SIC_i e $EIC_{(i)}$ foi proposta por Pregibon (1979, 1981) e Belsley, Khu e Welsch (1980). Estes autores partem de um esquema de perturbação (designado por perturbação infinitesimal) que consiste em alterar a ponderação dada a uma observação. Para o efeito, postulam que a variável residual tem variância σ^2 , à excepção da i -ésima observação, com $Var(u_i) = \sigma^2/w_i$ e $0 \leq w_i \leq 1$. De acordo com esta especificação, o estimador eficiente de β , estimador de Aitken, será

$$(2.32) \quad \beta(w_i) = (X^T W X)^{-1} X^T W Y,$$

em que W é uma matriz diagonal com o i -ésimo elemento igual a w_i e os restantes iguais à unidade. Aplicando o lema 1 a $(X^T W X)^{-1}$, o estimador (2.32) pode ser dado por [veja-se Belsley, Kuh e Welsch (1980) pag.65-66],

$$(2.33) \quad \hat{\beta}(w_i) = \hat{\beta} - \frac{(X^T X)^{-1} x_i^T \hat{u}_i (1 - w_i)}{1 - (1 - w_i) p_u}.$$

Para Pregibon (1981), o efeito de uma perturbação infinitesimal na variância da i -ésima variável residual, é dado pela diferenciação de (2.33) em relação a w_i , isto é,

$$(2.34) \quad \frac{d\hat{\beta}(w_i)}{dw_i} = \frac{(X^T X)^{-1} x_i^T \hat{u}_i}{[1 - (1 - w_i) p_u]^2}.$$

Então, à parte da dimensão da amostra, a EIC_i obtém-se calculando (2.34) em $w_i = 1$ e descreve a taxa de variação no estimador à medida que w_i se desvia de 1 isto é, à medida que se altera o peso dado à observação i . Do mesmo modo, a $EIC_{(w)}$ obtém-se de (2.34) quando $w_i \rightarrow 0$ e mede a taxa de variação no estimador resultante da omissão da i -ésima observação. A SIC_i pode ser vista como uma medida intermédia pois,

$$(2.35) \quad \begin{aligned} \frac{SIC_i}{n-1} &= \int_0^1 \frac{d\hat{\beta}(w_i)}{dw_i} dw_i \\ &= \hat{\beta}(1) - \hat{\beta}(0). \end{aligned}$$

2.3. MEDIDAS DA INFLUÊNCIA DE UMA OBSERVAÇÃO

Quando se está perante um modelo de regressão com $k > 1$ variáveis, a influência de cada observação no estimador, obtida a partir de uma das aproximações anteriores, pode ser vista como um ponto de \mathbb{R}^k . Estando-se particularmente interessado na ordenação desses pontos segundo o seu grau de influência, terá de se construir uma medida conveniente, isto é, uma aplicação de \mathbb{R}^k em \mathbb{R} . Uma possível medida é proposta por Hampel (1974) e dada por

$$(2.36) \quad \sup_z IC[z, T, F] \quad .$$

Hampel designa-a por '*Gross error sensitivity*' e pode ser vista como a maior influência, para qualquer z , exercida nos coeficientes estimados. A principal desvantagem de (2.36) está no facto de não ser invariante a modificações de localização e escala. Por outro lado, não se está tão interessado em encontrar a maior influência possível de uma observação nos resultados da regressão, mas apenas na sua ordenação em função do seu nível de influência. Uma via alternativa, implícita nos trabalhos de Cook (1977) e Cook e Weisberg (1982), passa pela construção de uma norma, determinada por uma matriz, M , simétrica, definida-positiva de ordem k e por um factor de escala, c , e conduz a,

$$(2.37) \quad D_i(M, c) = \frac{(IC[z_i, T, F])^T M (IC[z_i, T, F])}{c} \quad , \quad i = 1, 2, \dots, n \quad ,$$

em que a $IC[z_i, T, F]$ deve ser substituída por uma das aproximações anteriores e M e c devem escolher-se de modo a que $D_i(M, c)$ seja

invariante a alterações de escala e a transformações lineares não singulares das linhas da matriz X . Para uma escolha apropriada de M e c tem-se uma medida (escalar) da influência de cada uma das n observações.

A) Medida de Cook.

Num dos seus primeiros artigos, Cook (1977) propõe uma medida para avaliar o impacto de uma observação nos coeficientes estimados, sem qualquer referência à função de influência do estimador dos mínimos quadrados. Para tal, inspira-se na noção de elipsóide de confiança a $(1-\alpha).100\%$ para β , baseado em $\hat{\beta}$ e definido como o conjunto de todos os β^* tais que,

$$(2.38) \quad \frac{(\beta^* - \hat{\beta})^T X^T X (\beta^* - \hat{\beta})}{k \hat{\sigma}^2} \leq F_{1-\alpha}(k, n-k) .$$

Se o objectivo for o de avaliar a influência da observação i no vector dos coeficientes estimados, um procedimento natural passa, numa primeira fase, pelo cálculo de $\hat{\beta}$, sem a observação i , isto é, de $\hat{\beta}_{(i)}$. Substituindo posteriormente em (2.38), β^* por $\hat{\beta}_{(i)}$, a expressão que se obtém,

$$(2.39) \quad D_i = D_i(X^T X, k \hat{\sigma}^2) = \frac{(\hat{\beta}_{(i)} - \hat{\beta})^T X^T X (\hat{\beta}_{(i)} - \hat{\beta})}{k \hat{\sigma}^2} ,$$

constitui uma medida que pode ser interpretada como a distância quadrática entre $\hat{\beta}$ e $\hat{\beta}_{(i)}$, relativa à geometria fixa de $(X^T X)/k\hat{\sigma}^2$. Um valor elevado para D_i indica que a i -ésima observação tem um peso considerável na determinação de $\hat{\beta}$.

Só mais tarde, Cook e Weisberg (1980,1982) integram esta abordagem no âmbito da teoria das funções de influência. Tendo presente as várias aproximações anteriormente descritas, a medida de Cook pode ser construída a partir da *SIC*. Substituindo-se em (2.37) $IC(z_i, T, F)$ por $(n-1)SIC_i$, tem-se

$$(2.40) \quad D_i(M, c) = (n-1)^2 \frac{(SIC_i)^T M (SIC_i)}{c} \\ = \frac{(\hat{\beta}_{(i)} - \hat{\beta})^T M (\hat{\beta}_{(i)} - \hat{\beta})}{c}.$$

Se em (2.40) se fixar a métrica $M = X^T X$ e o factor de escala $c = k\hat{\sigma}^2$, chega-se à medida de influência de Cook, apresentada em (2.39).

A questão que obviamente se coloca é a de saber a partir de que valor de D_i , a observação i pode ser considerada como influente. Dada a similitude entre (2.38) e (2.39), Cook (1977) sugere que se compare D_i com os correspondentes valores assumidos pela distribuição $F_{1-\alpha}(k, n-k)$. Portanto, a omissão da observação i leva a uma alteração dos coeficientes estimados para o elipsóide de confiança a $(1-\alpha) \cdot 100\%$, baseado em n observações ou, dito de outra forma, a magnitude da distância entre $\hat{\beta}$ e $\hat{\beta}_{(i)}$ pode ser avaliada, encontrando o nível α do elipsóide centrado em $\hat{\beta}$ e que passa por $\hat{\beta}_{(i)}$. Note-se que (2.39) não tem uma distribuição F . Esta é apenas introduzida para converter D_i numa escala mais familiar. Cook e Weisberg (1982) e Montgomery

e Peck (1982) indicam que valores de D_i em torno ou maiores do que 1, correspondentes a uma deslocação de $\beta_{(i)}$ para o elipsóide de confiança acima de 50%, devem geralmente ser considerados como influentes. Por detrás desta regra não existe um rigoroso procedimento matemático. É em contrapartida uma regra prática resultante da aplicação desta medida a diferentes trabalhos empíricos.

A obtenção dos valores para D_i , $i=1,2,\dots,n$, obriga ao cálculo de $n+1$ regressões. Para problemas com uma dimensão razoável, tal facto pode levar a um grande dispêndio computacional. Neste sentido, é possível obter-se uma expressão equivalente, menos exigente em tempo de CPU e com interpretação mais sugestiva, em termos da informação proveniente da matriz de projecção e dos resíduos. Substituindo em (2.39) $\beta_{(i)}$ por (2.25) chega-se, após algumas simplificações ao seguinte,

$$(2.41) \quad D_i = \frac{1}{k} \frac{\hat{u}_i^2}{\hat{\sigma}^2(1-p_u)} \frac{p_u}{1-p_u} ,$$

$$= \frac{1}{k} t_i^2 \frac{p_u}{1-p_u} .$$

Então, à parte a constante k , D_i é o produto do quadrado dos resíduos studentizados (definido em (1.21)), pelo rácio $p_u/(1-p_u)$ que, de acordo com (1.11) pode ser visto como a distância de x_i ao centro do elipsóide definido por $(X_{(i)}^T X_{(i)})^{-1}$. Se o número de regressores, k , é fixo, a magnitude de D_i será determinada por dois factores: por um lado pelo valor de t_i , que por ser uma transformação monótona de r_i (definido em (1.23)), constitui uma medida para a detecção de

'outliers'; por outro, pela distância de x_i às restantes linhas da matriz X , reflectida por p_u , que é uma medida para a detecção de 'leverages'. Em síntese está-se agora em condições de poder afirmar que 'outliers' ou 'leverages' nem sempre correspondem a observações influentes⁹, enquanto que estas resultam necessariamente de uma combinação das duas primeiras.

Se o objectivo do investigador for o de avaliar a influência numa combinação linear, ou num subconjunto $q < k$, dos coeficientes, Cook (1977) propõe a utilização de uma norma de dimensão inferior, isto é, de uma matriz A tal que $r(A) = q < k$. Quando se está perante uma observação influente, este procedimento tem ainda a vantagem de permitir identificar o ou os coeficientes mais afectados.

Se L é uma matriz de ordem $(q \times k)$ com $r(L) = q < k$, tal que $\psi = L\beta$ constitua a combinação desejada, a SIC_t para $\hat{\psi} = L\hat{\beta}$ é, de modo similar a (2.29), dada por,

$$(2.42) \quad SIC_t = (n-1)(\hat{\psi} - \hat{\psi}_{(t)})^2, \quad t = 1, 2, \dots, n.$$

Substituindo $IC[z_i, T, F]$ em (2.37) por $(n-1)^{-1}SIC_t$, a influência da observação i sobre $\hat{\psi}$ obtém-se a partir da seguinte medida,

⁹ Em consequência é comum a designação de 'bons outliers' e 'bons leverages'. Os primeiros, são 'outliers' correspondentes a valores reduzidos de p_u e portanto a sua presença apenas influencia o termo autónomo do modelo. Em termos geométricos a sua presença pode ser vista como uma deslocação paralela do hiperplano definido por $Y = X\beta$. Os segundos são 'leverages' (p_u elevado) correspondentes a um valor quase nulo para os resíduos, não tendo quaisquer efeitos sobre os coeficientes do modelo.

$$(2.43) \quad D_i(L) = D_i(A, c) = \frac{(\hat{\psi} - \hat{\psi}_{(i)})^T A (\hat{\psi} - \hat{\psi}_{(i)})}{c},$$

em que, do mesmo modo, A e c devem ser escolhidos de forma a que $D_i(L)$ seja invariante a modificações de escala e a transformações lineares não singulares das linhas da matriz X . Para o efeito, Cook (1977) sugere que se tome $A = [L(X^T X)^{-1} L^T]^{-1}$ e $c = q\hat{\sigma}^2$. Nestas condições, fazendo $M = L^T A L$, (2.43) pode ser expressa de uma forma similar a (2.39) por,

$$(2.44) \quad D_i(L) = \frac{(\beta - \hat{\beta}_{(i)})^T M (\beta - \hat{\beta}_{(i)})}{q\hat{\sigma}^2},$$

$$= \frac{1}{q} \sum_i^2 \frac{x_i (X^T X)^{-1} M (X^T X)^{-1} x_i^T}{1 - p_{ii}}.$$

Sem perda de generalidade, admita-se uma situação concreta, em que se procura estudar o impacto nos últimos q elementos de β . Então, considerando as partições $X = [X_1 | X_2]$, com X_2 de ordem $(n \times q)$ e $L = [0 | I_q]$, onde I_q é a matriz identidade de ordem q , demonstra-se que,

$$(2.45) \quad x_i (X^T X)^{-1} M (X^T X)^{-1} x_i^T = x_i (X^T X)^{-1} x_i^T - x_i \begin{bmatrix} (X_1^T X_1)^{-1} & 0 \\ 0 & 0 \end{bmatrix} x_i^T.$$

Sendo $P_1 = X_1 (X_1^T X_1)^{-1} X_1^T$ a matriz de projecção para X_1 e $W = (I - P_1) X_2$ a projecção de X_2 no complemento ortogonal de X_1 , a expressão (2.45) simplifica-se para

$$(2.46) \quad x_i (X^T X)^{-1} M (X^T X)^{-1} x_i^T = p_{ii} - p_{1ii},$$

$$= w_{ii},$$

em que p_{iu} e w_u são, respectivamente, o i -ésimo elemento diagonal da matriz P_1 e da matriz $P - P_1 = P_2 = W(W^T W)^{-1} W^T$ [veja-se a propriedade 3]. Substituindo (2.46) em (2.44), tem-se finalmente,

$$(2.47) \quad D_i(L) = \frac{1}{q} t_i^2 \frac{w_u}{1 - p_u}.$$

Portanto, a influência do par (x_i, y_i) nos últimos q elementos de β , à parte a constante $1/q$, será tanto maior quanto maior t_i^2 e $w_u/(1 - p_u)$, em que a dimensão deste último é função directa da distancia da i -ésima linha de $(I - P_1)X_2$ à média amostral.

Quando se está em presença de uma observação influente pode ser particularmente interessante identificar o seu impacto em cada um dos coeficientes estimados. Se o objectivo for o de analisar o impacto no coeficiente $j=1, 2, \dots, k$, L será uma matriz linha de ordem $(1 \times k)$ onde o j -ésimo elemento é igual à unidade e os restantes nulos, e a expressão análoga a (2.47) pode ser dada por,

$$(2.48) \quad D_i(L) = k D_i(X^T X, k \hat{\sigma}^2) \text{Cor}^2(x_i \beta, L \beta),$$

onde $\text{Cor}^2(x_i \beta, L \beta)$ é o quadrado do coeficiente de correlação entre $x_i \beta = \hat{y}_i$ e $L \beta = \hat{\beta}_j$, isto é,

$$(2.49) \quad \text{Cor}^2(x_i \beta, L \beta) = \frac{[x_i (X^T X)^{-1}]_j^2}{p_u (X^T X)^{-1}_{jj}},$$

com $(X^T X)^{-1}_{.j}$ e $(X^T X)^{-1}_{jj}$ a coluna j e o j -ésimo elemento diagonal da matrix $(X^T X)^{-1}$.

Belsley, Kuh e Welsch (1980) chegam a um resultado similar ao estudarem o impacto da observação i sobre o coeficiente j , em que a medida desse impacto é dada pela diferença $\beta_j - \hat{\beta}_{(i)j}$ ¹⁰, relativa ao desvio padrão de $\hat{\beta}_j$, com σ estimado por $\hat{\sigma}_{(i)}$, isto é

$$(2.50) \quad DFBETAS_{ij} = \frac{\hat{\beta}_j - \hat{\beta}_{(i)j}}{[Var(\hat{\beta}_j)]^{1/2}} ,$$

$$= k[D_i(X^T X, k\hat{\sigma}_{(i)}^2)]^{1/2} Cor(x_i \hat{\beta}, L\hat{\beta}) .$$

Como $0 \leq Cor^2(.,.) \leq 1$, o impacto no coeficiente j , medido por (2.48), será sempre inferior ao impacto em todos os coeficientes. Cook (1977), mostra ainda que o valor máximo de $D_i(L)$, para i fixo, é atingido em $L = x_i$, isto é,

$$(2.51) \quad D_i(L) \leq D_i(x_i) = k D_i(X^T X, k\hat{\sigma}^2) .$$

Portanto, pode-se desde já concluir que se a influência de uma dada observação for diminuta, é desnecessário estudar a influência individual em cada coeficiente ou para qualquer combinação linear desses coeficientes.

B) Medida de Welsch-Kuh.

Welsch e Kuh (1977), apresentam uma medida para a influência da observação i em \hat{y}_i . Dada a relação (1.3), entre \hat{Y} e $\hat{\beta}$, ver-se-á que essa medida pode ser vista em termos da influência em $\hat{\beta}$.

10- Os autores atribuem a esta diferença a designação ' $DFBETA_j$ '.

A dedução desta medida de influência passa numa primeira fase, pela obtenção do valor estimado da i -ésima variável endógena sem a observação i , $\hat{y}_{i(i)} = x_i \hat{\beta}_{(i)}$ calculando numa segunda fase a diferença $\hat{y}_i - \hat{y}_{i(i)}$ ¹¹, estandarizada pelo desvio padrão de \hat{y}_i , isto é,

$$(2.52) \quad \frac{\hat{y}_i - \hat{y}_{i(i)}}{\sigma p_u^{1/2}}.$$

Para a estimação de σ estes autores e Belsley, Kuh e Welsch (1980) sugerem a utilização de $\hat{\sigma}_{(i)}$, definida em (1.22), pelo facto de esta ser robusta à observação i . Como $\hat{y}_i - \hat{y}_{i(i)} = x_i(\hat{\beta} - \hat{\beta}_{(i)})$ então, substituindo $\hat{\beta} - \hat{\beta}_{(i)}$ por (2.25), o resultado (2.52) pode ser dada por,

$$(2.53) \quad DFFITS_i = \left(\frac{p_u}{1 - p_u} \right)^{1/2} \frac{\hat{u}_i}{\hat{\sigma}_{(i)}(1 - p_u)^{1/2}},$$

ou, atendendo à expressão (1.23) para r_i por,

$$(2.54) \quad DFFITS_i = r_i \left(\frac{p_u}{1 - p_u} \right)^{1/2},$$

em que a terminologia 'DFFITS' é devida a Belsley, Kuh e Welsch. Comparando (2.54) com a medida de Cook (2.41), constata-se que a principal diferença está no factor de escala utilizado, pelo que

$$(2.55) \quad DFFITS_i = [D_i(X^T X, \sigma_{(i)})]^{1/2}$$

11- Belsley, Kuh e Welsch designam esta diferença por $DFFIT_i$.

Do mesmo modo, coloca-se a questão de saber qual o valor crítico a considerar. Apesar de (2.55) não ter uma distribuição t de Student, esta pode ser utilizada como um indicador, levando Velleman e Welsch (1981) a tomar como influentes, observações correspondentes a um valor de (2.55) superior a 1 ou 2.

Como r_i tem uma distribuição t de Student e atendendo a que os pontos correspondentes a $p_u = k/n$ podem ser identificados com uma situação ideal de não existência de 'leverages', um outro limite crítico possível é dado por,

$$(2.56) \quad t_{\alpha}(n-k-1) \left(\frac{k}{n-k} \right)^{1/2} .$$

Uma escolha razoável para t será 2, pelo que os valores de (2.54) deverão ser confrontados com¹²,

$$2 \left(\frac{k}{n-k} \right)^{1/2} .$$

C) Medida de Welsch.

Welsch (1982) parte da $EIC_{(i)}$, definida em (2.26), como uma aproximação à função de influência (2.21). Substituindo a expressão da EIC (2.26) em (2.37) obtém,

12- Belsley, Kuh e Welsch (1980), seguindo a mesma metodologia propõem de uma forma simplificada $2(k/n)^{1/2}$.

$$(2.57) \quad D_i(M, c) = \frac{(EIC_{(i)})^T M (EIC_{(i)})}{c}.$$

Na definição de M e c , Welsch procura que estes sejam robustos à observação i . Para o efeito, vai considerar $M = X_{(i)}^T X_{(i)} = (X^T X - x_i^T x_i)$ e $c = (n-1)\hat{\sigma}_{(i)}^2$, pelo que (2.57) pode ser dada pela seguinte expressão equivalente,

$$(2.58) \quad W_i = D_i(X_{(i)}^T X_{(i)}, (n-1)\hat{\sigma}_{(i)}^2) = (n-1)r_i^2 \frac{p_u}{(1-p_u)^2}.$$

Relacionando (2.58) com (2.54) tem-se,

$$W_i = (DFFITS_i)^2 \frac{n-1}{1-p_u},$$

onde se denota uma maior sensibilidade à presença de '*leverages*', atendendo a $1-p_u$ em denominador.

D) Medida de Atkinson.

Atkinson (1981), sugere que se utilize uma versão modificada da medida de Cook, propondo $\hat{\sigma}_{(i)}^2$ em vez de $\hat{\sigma}^2$ e tomando a raiz quadrada de D_i , corrigida pelos graus de liberdade. Esta modificação levou ao seguinte resultado:

(2.59)

$$\begin{aligned}
 T_i &= \left\{ \frac{n-k}{k} \frac{p_{ii}}{1-p_{ii}} \right\}^{1/2} |r_i|, \\
 &= | [D_i(X^T X, k/(n-k) \hat{\sigma}_{(i)}^2)]^{1/2} |, \\
 &= | DFFITS_i | \left(\frac{n-k}{k} \right)^{1/2}.
 \end{aligned}$$

As razões desta modificação devem-se essencialmente a dois factores: por um lado, T_i dá mais ênfase a observações extremas devido à consideração de $\hat{\sigma}_{(i)}^2$; por outro lado, no caso de não existirem 'leverages', $p_{ii} = k/n$, e portanto $T_i = |r_i|$, isto é, a influência da i -ésima observação depende apenas da existência de 'outliers' em Y .

Da análise anterior pode-se, a título de síntese, extrair duas importantes conclusões: por um lado, as medidas de influência apresentadas podem ser deduzidas a partir de uma das diferentes aproximações feitas à função de influência teórica; por outro, é possível para uma mesma aproximação construir diferentes medidas, consoante a métrica e o factor de escala utilizados em (2.37).

A excepção da medida de Welsch (2.58), todas as restantes podem ser derivadas a partir da SIC (ou da SC). As razões subjacentes à escolha desta aproximação pelos diferentes autores devem-se, como se frisou no final da secção 2.2, ao facto de poderem ser interpretadas como a variação ocorrida nos coeficientes estimados quando se omite uma observação, e por constituírem medidas intermédias entre a EIC_i e a $EIC_{(i)}$.

Utilizando a *SIC* como aproximação à função de influência (2.13), as medidas de Cook, Welsch-Kuh e Atkinson distinguem-se essencialmente pelo factor de escala utilizado. A medida de Cook, como uma medida da influência em $\hat{\beta}$, é dada pela distância, $\hat{\beta} - \hat{\beta}_{(i)}$, relativa à geometria fixa de $X^T X / \hat{\sigma}^2$. Para as medidas de Welsch-Kuh e Atkinson o mesmo tipo de raciocínio deixa de ser válido, pois a regra utilizada para comparar $\hat{\beta}$ e $\hat{\beta}_{(i)}$, isto é, $X^T X / \hat{\sigma}_{(i)}^2$, varia de observação para observação, dificultando a sua interpretação como medidas da influência em $\hat{\beta}$. Em consequência, estas medidas devem ser interpretadas, de acordo com Welsch (1982) e Chatterjee e Hadi (1988), como medidas da influência, simultaneamente em $\hat{\beta}$ e $\hat{\sigma}^2$, relativa a $X^T X$. O mesmo tipo de conclusão pode ser tirado para a medida de Welsch. A principal diferença está no facto de esta ser definida relativamente a $X_{(i)}^T X_{(i)}$, que não é fixa, dificultando ainda mais a sua interpretação.

2.4. MEDIDAS DA INFLUÊNCIA DE UM SUBCONJUNTO DE OBSERVAÇÕES.

A razão desta secção prende-se com o facto de certas observações só serem influentes quando consideradas conjuntamente com outra ou outras observações. Esta situação é também referida por 'efeito de disfarce' (*masking effect*, na literatura inglesa)¹³ em virtude de a influência de uma dada observação estar disfarçada

13- Esta designação é devida a Atkinson (1986).

pela presença de outra. Em consequência, a aplicação de uma das medidas (2.41), (2.54), (2.58) ou (2.59) torna-se ineficaz para a sua detecção.

As diferentes medidas de influência de um subconjunto $m < k$ das observações podem ser vistas como uma generalização das medidas apresentadas na secção 2.3. Para o efeito, vai-se admitir que $I^T = (i_1, i_2, \dots, i_m)$, $1 \leq i_j \leq n$, é um vector de índices, relativo a m observações, cuja influência se procura conhecer. No que se segue, e em estreita analogia com a notação anterior, os símbolos ' I ' e ' I ' em índice inferior de um vector ou matriz representam, respectivamente, a omissão de m observações indexadas por I e o vector ou matriz com as observações definidas por I . Sendo assim, as medidas de Cook, Welsch-Kuh, Welsch e Atkinson para m observações são em termos formais idênticas às que se apresentaram na secção anterior, bastando que se substituam os índices em i por I . Deste modo, a expressão equivalente a (2.40), resultante da substituição da SIC_i pela SIC_I em (2.37), e que serve de suporte para a derivação das medidas de Cook, Welsch-Kuh e Atkinson, vem dada por,

$$(2.60) \quad D_I(M, c) = \frac{(\hat{\beta}_{(I)} - \hat{\beta})^T M (\hat{\beta}_{(I)} - \hat{\beta})}{c}.$$

Esta igualdade pode também ser apresentada em termos dos resíduos studentizados e dos elementos diagonais da matriz P . Para tal vai-se expressar $\hat{\beta}_{(I)}$ por,

$$(2.61) \quad \begin{aligned} \hat{\beta}_{(I)} &= (X_{(I)}^T X_{(I)})^{-1} X_{(I)}^T Y_{(I)} \\ &= (X^T X - X_I^T X_I)^{-1} (X^T Y - X_I^T Y_I). \end{aligned}$$

Aplicando posteriormente o lema 1 ao primeiro factor do segundo membro de (2.61), e definindo P_I como uma submatriz de P , obtida pelo cruzamento das m linhas e colunas indexadas por I , chega-se após algumas simplificações algébricas ao seguinte,

$$(2.62) \quad \hat{\beta}_{(I)} = \hat{\beta} - (X^T X)^{-1} X_I^T [(I + (I - P_I)^{-1} P_I) Y_I - (I - P_I)^{-1} X_I \hat{\beta}] .$$

Como $(I - P_I)^{-1} = I + (I - P_I)^{-1} P_I$ [veja-se Cook e Weisberg (1982), pág. 136]

(2.62) simplifica-se para,

$$(2.63) \quad \hat{\beta}_{(I)} = \hat{\beta} - (X^T X)^{-1} X_I^T (I - P_I)^{-1} \hat{u}_I .$$

Substituindo (2.63) em (2.60) tem-se,

$$(2.64) \quad D_I(M, c) = \frac{\hat{u}_I^T (I - P_I)^{-1} X_I (X^T X)^{-1} M (X^T X)^{-1} X_I^T (I - P_I)^{-1} \hat{u}_I}{c} .$$

Nestas condições a medida de Cook para m observações, apresentada por Cook e Weisberg (1980), obtém-se de (2.64) para $M = X^T X$ e $c = k \hat{\sigma}^2$, isto é,

$$(2.65) \quad D_I = D_I(X^T X, k \hat{\sigma}) = \frac{\hat{u}_I^T (I - P_I)^{-1} P_I (I - P_I)^{-1} \hat{u}_I}{k \hat{\sigma}^2} .$$

Um valor elevado para (2.65) indica que as observações indexadas por I são conjuntamente influentes em $\hat{\beta}$. No caso particular em que $m=2$, isto é, $I=(r,s)$, Gray e Ling (1984) apresentam (2.65) por,

$$(2.66) \quad k \hat{\sigma}^2 D_I = k \hat{\sigma}^2 (D_r + D_s) \left(1 + \frac{p_{rs}^2}{d_{rs}^2} \right)^2 + p_{rs}^2 \frac{[\hat{u}_r^2 (2 - p_{rr}) + \hat{u}_s^2 (2 - p_{ss})]}{d_{rs}^2} + \frac{2 \hat{u}_r \hat{u}_s p_{rs} (1 + p_{rs}^2 - p_{rr} p_{ss})}{d_{rs}^2} ,$$

onde,

$$d_{rs} = (1 - p_{rr})(1 - p_{ss}) - p_{rs}^2 .$$

Como pela propriedade 5(a) $d_{rs} \geq 0$, se $(\hat{u}_r \hat{u}_s p_{rs}) > 0$, então $D_i > D_r + D_s$, isto é, a influência conjunta é sempre superior à soma das influências individuais. A igualdade é verificada se a matriz P_i for diagonal ($p_{rs} = 0, \forall r \neq s$).

No caso mais geral em que I tem dimensão m , pode-se em alternativa a (2.65) derivar uma expressão mais conveniente, mediante a decomposição da matriz P_i em,

$$(2.67) \quad P_i = \Gamma \Pi \Gamma^T ,$$

onde Π é a matriz diagonal dos valores próprios, $0 \leq \pi_1 \leq \pi_2 \leq \dots \leq \pi_m \leq 1$, e $\Gamma = [\gamma_1, \gamma_2, \dots, \gamma_m]$ a matriz dos vectores próprios associados. Substituindo (2.67) em (2.65) obtém-se,

$$(2.68) \quad D_i = \frac{(\Gamma^T \hat{u}_i)^T (I - \Pi)^{-1} \Pi (I - \Pi)^{-1} (\Gamma^T \hat{u}_i)}{k \hat{\sigma}^2} .$$

Para que esta expressão tenha sentido π_m tem de ser diferente de 1; caso contrário a matriz $(I - \Pi)$ será singular. Neste último caso, Cook e Weisberg (1980, 1982) consideram $D_i = \infty$.

Atendendo a que,

$$\Gamma^T \hat{u}_i = [\gamma_1^T \hat{u}_i \quad \dots \quad \gamma_m^T \hat{u}_i]^T ,$$

e

$$\frac{1}{k \hat{\sigma}^2} = \frac{1}{n} \frac{1}{\hat{\sigma}^2} ,$$

$$(2.73) \quad (I - \Pi)^{-1} \Pi (I - \Pi)^{-1} = \begin{bmatrix} \frac{\pi_1}{(1 - \pi_1)^2} & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & \frac{\pi_m}{(1 - \pi_m)^2} \end{bmatrix},$$

tem-se,

$$(2.69) \quad D_I = \frac{\sum_{j=1}^m (\mathbf{y}_j^T \hat{\mathbf{u}}_I)^2 \frac{\pi_j}{(1 - \pi_j)^2}}{k \hat{\sigma}^2}.$$

Fazendo, de acordo com Cook e Weisberg (1980),

$$(2.70) \quad h_j^2 = \frac{(\mathbf{y}_j^T \hat{\mathbf{u}}_I)^2}{\hat{\sigma}^2 (1 - \pi_j)}, \quad j = 1, 2, \dots, m,$$

a medida de Cook para m observações pode ser dada em analogia com

(2.41) por,

$$(2.71) \quad D_I = \frac{1}{k} \sum_{j=1}^m h_j^2 \frac{\pi_j}{1 - \pi_j},$$

em que o papel de t_i^2 é agora desempenhado por h_j^2 e $p_u/(1 - p_u)$ é substituído por $\pi_j/(1 - \pi_j)$.

Esta analogia carece no entanto de alguns comentários prévios. A generalização do quadrado dos resíduos studentizados (1.21), como uma medida para a detecção de m 'outliers', é de acordo com Gentleman e Wilk (1975), dada por,

$$(2.72) \quad \sum_{j=1}^m h_j^2 = \frac{\hat{\mathbf{u}}_I^T (I - P_I) \hat{\mathbf{u}}_I}{\hat{\sigma}^2} = t_I^2,$$

onde, sob normalidade os h_j^2 são independentes. Na hipótese nula de as m observações não serem 'outliers', a estatística,

$$(2.73) \quad F_I = \frac{(n-k-m) \sum h_j^2}{n-k - \sum h_j^2},$$

tem uma distribuição F com m e $n-k-m$ graus de liberdade.

A generalização da medida para a detecção de m 'leverages' é menos pacífica. Em estreita analogia com $p_u/(1-p_u)$, Cook e Weisberg (1980) propõem,

$$(2.74) \quad \sum_{j=1}^m \frac{\pi_j}{1-\pi_j} = \text{tr}[P_I(I-P_I)^{-1}].$$

Para o caso $m=1$, viu-se no capítulo 1 que p_u constituía uma medida adequada para a detecção de 'leverages', tal como qualquer transformação monótona deste, como é o caso de $p_u/(1-p_u)$. Para o caso $m>1$, o problema prende-se com o facto de (2.74) não ser uma transformação monótona de $\text{tr}(P_I)$. Nestas condições, Cook e Weisberg (1982) propõem uma factorização alternativa para (2.71). Multiplicando e dividindo (2.71) por $\sum h_j^2 = t_I^2$ obtém-se,

$$(2.75) \quad D_I = \frac{t_I^2}{k} \sum_{j=1}^m \left(\frac{h_j^2}{t_I^2} \right) \frac{\pi_j}{1-\pi_j}.$$

Definindo,

$$(2.76) \quad Q_I = \sum_{j=1}^m \left(\frac{h_j^2}{t_I^2} \right) \frac{\pi_j}{(1-\pi_j)},$$

como o 'potencial generalizado' (Cook e Weisberg (1982), pág.138),

(2.75) pode ser apresentada por,

$$(2.77) \quad D_I = \frac{t_I^2}{k} Q_I \quad ,$$

em que a influência conjunta das m observações indexadas por I é dada como uma combinação de t_I , por um factor, Q_I , que pode ser visto como uma medida para a detecção de 'leverages'. A partir da desigualdade ¹⁴,

$$(2.78) \quad \frac{\pi_1}{1-\pi_1} \leq Q_I \leq \frac{\pi_m}{1-\pi_m} \quad ,$$

a presença de 'leverages' em X pode ser averiguada pela magnitude de π_1 e π_m ou, atendendo a que $\pi_m \leq tr(P_I) = \sum_{i \in I} p_{ii}$ e $\pi_1 \leq \min(p_{ii})$, pela magnitude de p_{ii} , com $i \in I$. Sendo assim, se p_{ii} é pequeno também o será π_1 e Q_I . Por outro lado, se p_{ii} é elevado também o será π_1 e portanto Q_I .

Para a derivação das medidas de Welsch-Kuh e Atkinson procede-se de modo análogo, realçando apenas que $\hat{\sigma}_{(I)}$ de (2.54) e (2.59) deva ser substituído por $\hat{\sigma}_{(I)}$ em que,

$$(2.79) \quad \hat{\sigma}_{(I)}^2 = \hat{\sigma}^2 \left(\frac{n-k-t_I}{n-k-m} \right) \quad ,$$

com t_I definido em (2.72).

14- Note-se que, ao contrário do potencial definido para uma observação, Q_I tem uma componente aleatória. No entanto, esta componente é independente dos parâmetros do modelo e o intervalo apresentado é não estocástico.

A medida de Welsch obtém-se de (2.37), substituindo a $IC[z_i, T, F]$ pela EIC baseada em $(n-m)$ observações e M e c por $X_{(i)}^T X_{(i)}$ e $(n-m)\hat{\sigma}_{(i)}^2$, respectivamente. Estas levam ao seguinte,

$$(2.80) \quad D_i[X_{(i)}^T X_{(i)}, (n-m)\hat{\sigma}_{(i)}^2] = \frac{(n-m)}{\hat{\sigma}_{(i)}^2} \hat{u}_i^T (I - P_i)^{-2} P_i (I - P_i)^{-1} \hat{u}_i \quad .$$

Utilizando a decomposição de P_i em valores e vectores próprios prova-se, por um processo análogo ao que se descreveu para a medida de Cook que a expressão anterior simplifica-se para,

$$(2.81) \quad D_i[X_{(i)}^T X_{(i)}, (n-m)\hat{\sigma}_{(i)}^2] = (n-m) \sum_{j=1}^m \frac{(\mathbf{y}_j^T \hat{\mathbf{u}}_i)^2}{\hat{\sigma}_{(i)}^2 (1 - \pi_j)} \frac{\pi_j}{(1 - \pi_j)^2} \quad ,$$

Poderia a este nível pensar-se que o problema da influência estaria resolvido. Quer para o caso individual, $m=1$, quer para o caso múltiplo, a aplicação de uma das medidas anteriores permite lançar alguma luz sobre a sua influência. No entanto, o investigador desconhece a priori quais os subconjuntos de observações influentes. Estudar todos os subconjuntos possíveis constitui uma tarefa penosa pela dimensão dos cálculos a efectuar. Por exemplo, para $n=30$ e $m=5$ ter-se-iam 174436 subconjuntos distintos.

As tentativas de contornar este problema tem estado na origem do desenvolvimento da aplicação de métodos robustos, para a detecção de subconjuntos de observações influentes. Estes métodos, apresentados por Hawkins, Bradu e Kass (1984), Atkinson (1986) e Rousseeuw e Leroy (1987), fundamentam-se em traços gerais num

estimador que se obtém pela minimização da mediana do quadrado dos resíduos (*Least Median Squares*), proposto por Rousseeuw (1984). Sem o devido rigor matemático, pode-se afirmar que o fundamental destes métodos baseiam-se na 'identificação' das 'boas' observações, isto é, daquelas observações que não sejam consideradas 'outliers' nem 'leverages', efectuando posteriormente uma regressão sobre estas observações. Dado que as 'más' observações não foram consideradas no processo de regressão, estas são facilmente identificadas por uma análise dos resíduos.

Em alternativa aos métodos robustos (que não serão tratados neste trabalho), foram desenvolvidas por Cook e Weisberg (1980) e Gray e Ling (1984) duas abordagens que procuram reduzir o número de subconjuntos a estudar.

A) Abordagem de Cook e Weisberg.

Estes autores procuram majorar D_j utilizando apenas a informação proveniente da matriz de projecção e dos resíduos. Em função desta informação, um determinado subconjunto será ou não seleccionado como um potencial influente.

Como por definição $\pi_m \geq \pi_j$, $j=1,2,\dots,m$, também $\pi_m/(1-\pi_m) \geq \pi_j/(1-\pi_j)$, em virtude de este ser uma transformação monótona daquele. Deste modo um limite superior para D_j , poderá ser dado por,

$$(2.82) \quad D_I \leq \frac{1}{k\hat{\sigma}_{(I)}^2(1-\pi_m)^2} \sum_{j=1}^m (\gamma_j^T \hat{u}_I)^2,$$

ou atendendo a que $\sum (\gamma_j^T \hat{u}_I)^2 = \hat{u}_I^T \Gamma \Gamma^T \hat{u}_I = \hat{u}_I^T \hat{u}_I$,

$$(2.83) \quad D_I \leq \frac{\pi_m}{(1-\pi_m)^2} \left(\frac{\sum_{j=1}^m \hat{u}_I^2}{k\hat{\sigma}^2} \right).$$

Para que o cálculo do membro direito da desigualdade (2.83) seja operacional, no sentido em que não seja necessário formar a matriz P_I , Cook e Weisberg propõem um substituto para π_m . Atendendo a que $\pi_m \leq \text{tr}(P_I)$ e desde que $\text{tr}(P_I) < 1$, a sua substituição em (2.83) leva ao seguinte,

$$(2.84) \quad D_I \leq \frac{\sum_{j \in I} p_{II}}{\left(1 - \sum_{j \in I} p_{II}\right)^2} \left(\frac{\sum_{j \in I} \hat{u}_I^2}{k\hat{\sigma}^2} \right).$$

Se $\text{tr}(P_I) \geq 1$, a sua substituição em (2.83) deixa de ser possível. Para m pequeno (não superior a três), os autores consideram que não se ganha em aproximar π_m , propondo que se aplique directamente (2.77) a cada subconjunto.

A aplicação deste método é eficiente se a relação entre o número de observações e o número de variáveis explicativas é elevada, em virtude de não haver muitos subconjuntos de observações correspondentes a $\text{tr}(P_I) \geq 1$.

B) Abordagem de Gray e Ling.

Gray e Ling (1984) propõem a aplicação de um método de

'cluster' à matriz P_z , para a detecção de potenciais subconjuntos de observações influentes com o principal objectivo de reduzir o número de subconjuntos a processar.

Dado que a matriz de projecção P_z , apresentada em (1.34), pode ser decomposta na soma de uma matriz P por uma outra que depende exclusivamente dos resíduos, esta contém segundo os autores toda a informação necessária para o estudo da influência. Em consequência, subconjuntos de observações potencialmente influentes podem ser associadas a submatrizes de P_z contendo valores elevados em magnitude, isto é, a valores que ultrapassem um certo limiar previamente fixado. Para uma melhor identificação destas submatrizes, sugere-se troca de linhas e colunas de P_z , de modo a que se obtenha uma matriz (quase) diagonal por blocos. De uma forma mais eficiente, os autores servem-se do algoritmo 'k-clustering' desenvolvido por Ling (1972).

O papel desempenhado pelos elementos de P_z , bem como a relação destes com a medida de Cook, é mais intuitiva no caso bidimensional, $I=\{r,s\}$, a partir da expressão (2.66). Por esta, tem-se que quanto menor d_{rs} e quanto maior $\hat{u}_r\hat{u}_s p_{rs}$, maior será o valor de D_I . No mesmo sentido, se $\hat{u}_r\hat{u}_s > 0$ e $d_{rs} > 0$, D_I é função crescente de p_{rs} ; se $\hat{u}_r\hat{u}_s < 0$ e $p_{rs} < 0$ D_I é função decrescente de p_{rs} , em que $-0.5 \leq p_{rs} \leq 0.5$.

Após se terem identificado estes subconjuntos potencialmente influentes, a aplicação de uma das medidas estudadas permite identificar quais os influentes.

3. ESTUDO DA INFLUÊNCIA NUM MODELO PARA A PROCURA DE MOEDA EM PORTUGAL.

Neste capítulo são aplicadas algumas medidas de influência a um modelo econométrico para a procura de moeda, M1⁻ (moeda em circulação fora das instituições monetárias mais depósitos à ordem, excluindo os do Sector Público), em Portugal, no período 1961-1990. Os objectivos subjacentes à elaboração desta aplicação foram os seguintes:

- avaliar sobre o comportamento das diferentes medidas da influência apresentadas e,
- identificar ao longo do período de estimação alguns dos factores que possam estar na origem de eventuais observações influentes.

3.1. ALGUMAS CONSIDERAÇÕES TEÓRICAS SOBRE A DEDUÇÃO DA FUNÇÃO PROCURA DE MOEDA.

Antes da apresentação dos resultados da estimação fazem-se algumas considerações teóricas sobre a procura de moeda no sentido de justificar o modelo que irá servir de suporte a este capítulo. Saliente-se que um estudo exaustivo da procura de moeda em termos teóricos está fora dos objectivos deste trabalho.

O estudo da procura de moeda neste século, em especial após os anos 20/30, tem sofrido um desenvolvimento divergente, consoante as diferentes correntes económicas. A parte do confronto entre as duas principais Escolas Teóricas (Monetaristas e Keynesianos), é no entanto possível obter-se algum consenso no que respeita às variáveis explicativas do modelo. Em traços gerais, a procura de moeda em termos reais é função de uma variável de escala (em geral, do rendimento real, do rendimento permanente ou da riqueza dos agentes) e de uma outra que mede o custo de oportunidade da sua detenção. Designando por Y a variável de escala e R^1 o custo de oportunidade tem-se,

$$(3.1) \quad \frac{M_t^d}{P_t} = f(Y_t, R_t) \quad ,$$

em que o índice i é o tempo e M^d/P é a procura por saldos reais, com M^d , os saldos nominais e P o nível dos preços. A equação (3.1) tem implícita a hipótese de a elasticidade moeda-preços ser unitária. Para alguns autores, entre os quais Goodhart, C.A.E. (1989), esta hipótese deve ser devidamente testada, pois nem sempre é validada pelos estudos empíricos. Neste sentido, optou-se pela reformulação de (3.1) para,

$$(3.2) \quad M_t^d = g(P_t, Y_t, R_t) \quad ,$$

1 - O custo de oportunidade é em geral apresentado por um vector de variáveis. As que com maior frequência aparecem nos estudos empíricos são a taxa de juro de curto e longo prazo. Como tal, daqui em diante vai-se atribuir a designação genérica de taxa de juro a R .

ou, admitindo uma especificação log-linear², para,

$$(3.3) \quad m_t^d = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \alpha_3 r_t,$$

onde m^d , p , y e r representam o logaritmo natural de M^d , P , Y e R , respectivamente. De acordo com a teoria econômica, espera-se que a procura de moeda em termos nominais varie positivamente com o nível dos preços, positivamente com o rendimento real e inversamente com a taxa de juro.

A equação (3.3) deve ser interpretada como uma relação de equilíbrio de longo prazo, isto é, como uma relação de equilíbrio onde a quantidade de moeda em circulação (oferta de moeda) e o nível de procura desejado pelos agentes económicos se igualem. Havendo, por exemplo, uma variação da oferta de moeda, a passagem de uma situação de equilíbrio a outra é estática. O que entretanto se passa, ou a forma como os agentes económicos se ajustam à nova situação de equilíbrio, não é contemplado por este modelo. Em consequência, dado que as estatísticas monetárias fornecem a quantidade de moeda em circulação e como, em cada momento, esta não tem necessariamente que coincidir com a procura desejada pelos

2 - A opção por uma especificação log-linear, tradicional neste tipo de modelos, visa essencialmente facilitar a interpretação dos coeficientes das variáveis exógenas do modelo. Sem entrar nesse campo, refira-se no entanto que a escolha por uma especificação linear ou log-linear é passível de ser testada.

agentes³, a aplicação directa dos Mínimos Quadrados a (3.3), proporciona resultados que em geral não são satisfatórios⁴.

Pelos motivos apontados, a estimação de uma função procura de moeda, requer que se entre em linha de conta com a forma como os agentes se ajustam ao nível desejado de moeda. Para o efeito existem na literatura dois tipos de ajustamentos: (i) um parcial, devido a Chow (1966) e (ii) outro mais geral mediante uma modelização tipo ADL (Autoregressive Distributed Lag), proposto inicialmente por Davidson, Hendry, Sriv e Yeo (1978) para a função consumo. Apesar de (i) poder ser visto como um caso particular de (ii)⁵ vai-se, por uma questão de simplicidade, propor um esquema de ajustamento parcial para a procura de moeda.

De acordo com (i), a variação do stock de moeda detido pelos agentes, num determinado momento, deverá ser igual a uma proporção da diferença entre o nível desejado e o nível verificado no momento anterior, isto é,

$$m_t - m_{t-1} = \theta(m_t^d - m_{t-1})$$

ou,

$$(3.4) \quad m_t = \theta m_t^d + (1 - \theta)m_{t-1}$$

3 - Se se admitir que os agentes não detêm toda a informação disponível (informação incompleta) e se existirem custos associados à alteração do stock de activos, o que é razoável.

4 - Em particular, os resíduos estimados do modelo não são independentes.

5 - Veja-se por exemplo Cuthbertson e Taylor (1987), pág. 66.

Substituindo m_t^d de (3.3) em (3.4), a função procura de moeda de curto prazo vem dada por,

$$(3.5) \quad m_t = \beta_0 + \beta_1 p_t + \beta_2 y_t + \beta_3 r_t + \beta_4 m_{t-1},$$

com $\beta_j = \theta \alpha_j$, $j = 0, 1, 2, 3$ e $\beta_4 = 1 - \theta$.

3.2. RESULTADOS DA ESTIMAÇÃO DO MODELO PARA A PROCURA DE MOEDA.

De acordo com a terminologia utilizada em Portugal para representar os agregados monetários, a série relativa à procura de moeda, corresponde ao agregado $M1^-$ (moeda em circulação fora das instituições monetárias mais depósitos à ordem, excluindo os do sector público). As restantes séries têm o seguinte significado: p - deflactor do consumo privado com base em 1977; y - PIBpm a preços constantes de 1977; r - taxa de juro dos depósitos a prazo de 6 meses a 12 meses.

Para a estimação do modelo (3.5), obtenção dos valores das estatísticas teste e cálculo das diferentes medidas da influência (secção 3.3), utilizou-se o programa '386 TSP vers. 4.2A' num computador AT 386/387-25Mhz.

A estimação do modelo (3.5) pelo Método Ordinário dos Mínimos Quadrados (MOMQ), para os anos de 1961 a 1990, conduziu aos seguintes resultados,

QUADRO 3.1 - ESTIMATIVAS DO MOMQ

coef.	estimat.	desv.pad.	estat. t
β_0	-2.772	0.8051	-3.4432
β_1	0.3726	0.1047	3.5586
β_2	0.9748	0.2444	3.9878
β_3	-0.0932	0.0380	-2.4552
β_4	0.4294	0.1530	2.8077

N=30	$R^2 = 0.9981$	$\bar{R}^2 = 0.9979$
$F_{(4, 26)} = 3367.4$	$\hat{\sigma}^2 = 0.0582$	

De acordo com os resultados obtidos, as estimativas dos parâmetros têm o sinal esperado, são estatisticamente significativas (atendendo aos valores do rácio t) e, pela estatística F, as variáveis explicativas, no seu conjunto, são determinantes para a explicação da procura de moeda.

Tendo por base a proposta de Hendry (1980), sobre a necessidade de uma avaliação estatística mais completa do modelo, procedeu-se

à aplicação de um conjunto de testes, no sentido de detectar eventuais perturbações que possam afectar a eficiência do MOMQ⁶. De uma forma breve, os testes utilizados foram os seguintes:

DURBIN 70 - Teste de autocorrelação de 1ª ordem (AR(1) ou MA(1)) para modelos com variável endógena desfasada (D70). Esta estatística tem uma distribuição assintótica $t_{(n-k-2)}$. Refira-se ainda que a sua potência é afectada em amostras de dimensão reduzida.

HARVEY - Teste de autocorrelação de ordem q (AR(q) ou MA(q)). Este teste, apresentado por Harvey (1982), pode ser visto como uma generalização do teste anterior, sendo recomendado por Kiviet (1986) para amostras de pequena dimensão; tem distribuição $F_{(q, n-k-q)}$.

WHITE - Teste de heterocedasticidade, apresentado por White (1980), que resulta do confronto do estimador usual da matriz de variâncias-covariâncias de β com um outro estimador da mesma matriz robusto à heterocedasticidade. Tem a vantagem de não exigir nenhum conhecimento prévio quanto ao tipo de heterocedasticidade potencialmente presente e permite também inferir sobre a especificação do modelo.

ENGLE - Teste de heterocedasticidade construído a partir dos processos ARCH (heterocedasticidade condicional autoregressiva),

6 - Refira-se que a aplicação de testes de especificação a modelos dinâmicos só é, em geral, válida em termos assintóticos, sendo a sua potência fortemente afectada em pequenas amostras. Tendo por base os trabalhos de simulação de Kiviet (1985, 1986), seleccionaram-se, de entre os vários testes de diagnóstico, aqueles que melhor se têm revelado em pequenas amostras.

devido a Engle (1982), que ensaia a hipótese de homocedasticidade dos resíduos contra a alternativa de estes seguirem um processo ARCH(q).

Em traços gerais, um processo ARCH(q) caracteriza-se por,

$$U_t | \Omega_{t-1} \sim N(0, h_t) ,$$

$$h_t = h(U_{t-1}, \dots, U_{t-q}, \bar{\alpha}) ,$$

onde Ω_{t-1} é o conjunto de informação disponível no período t-1 e $\bar{\alpha}$ um vector de parâmetros desconhecidos.

A construção deste teste processa-se em duas fases,

- estimação da regressão auxiliar de \hat{U}_t^2 sobre $\hat{U}_{t-1}^2, \dots, \hat{U}_{t-q}^2$ mais uma constante e,
- cálculo de nR^2 , onde R^2 é o coeficiente de correlação múltipla da regressão anterior.

A estatística do teste é equivalente a nR^2 com distribuição assintótica $\chi^2_{[k(k+1)/2]}$.

CHOW - Teste de permanência de estrutura proposto por Chow (1960) e recomendado por Kiviet (1986), com distribuição $F_{(n_2, n_2-k)}$, onde n_1 é a dimensão amostral e n_2 a dimensão pós amostral.

Atendendo a que este teste é frequentemente utilizado sem a construção de uma particular hipótese alternativa⁷ (H_1), pode ser interpretado não só para a detecção de eventuais variações do valor

7 - Veja-se Kiviet (1986), pág. 248.

dos parâmetros, mas também noutros erros de especificação.

RINBOW - Teste à não linearidade do modelo e outros tipos de má especificação da forma funcional, proposto por Jessica Uts e referido em Kramer e Sonnberger (1986). Este teste pode ser visto como uma variante do teste de Chow, em virtude de confrontar a soma do quadrado dos resíduos do modelo estimado a partir de n observações com a do modelo estimado a partir de uma subamostra de dimensão $n/2$ formado por observações mais próximas do 'centro de gravidade' das variáveis explicativas, isto é, com um menor valor de p_{it} , $i=1,2,\dots,n$.

A estatística tem distribuição $F_{(n-n_1, n_1-k)}$.

BJ - Teste à normalidade dos resíduos apresentado por Bera e Jarque. Em traços gerais este teste mede o afastamento das estimativas de μ_3 e μ_4 de 0 e $3\sigma^4$, respectivamente. Sob a hipótese nula de os erros terem distribuição normal, a estatística do teste BJ tem uma distribuição assintótica $\chi^2_{(2)}$.

O cálculo das estatísticas que sumariamente se acabam de apresentar conduziu aos seguintes resultados,

3 - Centro de gravidade calculado de os dados originais a 25.

4 - A distribuição 20121 e 20122 refere-se ao teste de 50121 e 50122, respectivamente, com os dados originais e os dados transformados, respectivamente.

10 - A distribuição 20121 e 20122 refere-se ao teste de 50121 e 50122, respectivamente, com os dados originais e os dados transformados, respectivamente.

QUADRO 2 - TESTES ESTATISTICOS⁸

D70	HARVEY	WHITE	ENGLE1 ⁹
-0.0181 (-1.714)	0.0077 (3.42)	8.226 (25.0)	0.616 (3.841)
ENGLE2	RINBOW	CHOW(73) ¹⁰	CHOW(74)
0.995 (5.991)	0.743 (2.84)	2.352 (2.48)	1.209 (2.58)
CHOW(75)	CHOW(76)	CHOW(77)	BJ
1.389 (2.70)	0.998 (2.84)	0.950 (3.02)	1.405 (5.99)

Da análise destes resultados pode-se concluir que o modelo estimado é aceitavel, em virtude de ter 'passado' em todos os testes estatisticos que aqui se apresentam.

8 - Dentro de parêntesis encontram-se os níveis críticos a 5%.

9 - A designação ENGLE1 e ENGLE2 refere-se ao teste de ENGLE sob a hipótese alternativa de os resíduos seguirem um processo ARCH(1) e ARCH(2), respectivamente.

10 - A designação CHOW(i), $i=73, \dots, 77$, diz respeito ao teste de Chow em i. A selecção destes anos prende-se essencialmente com 1º choque petrolífero, 1973, e com a revolução de 1974.

3.3. IDENTIFICAÇÃO DAS OBSERVAÇÕES INFLUENTES.

A) Influência individual.

Nesta subsecção vão ser aplicadas as medidas de Cook (2.41), Welsch-Kuh (2.54), Welsch (2.58) e Atkinson (2.59) às observações do modelo estimado em 3.2. Atendendo a que estas medidas dependem essencialmente de uma transformação dos resíduos estimados, \hat{U} , e dos elementos diagonais da matriz de projecção, P , serão também apresentados os valores dos resíduos studentizados e dos elementos diagonais de P , no sentido de se conseguir averiguar as possíveis causas de uma eventual observação influente, isto é, 'Outlier' ou 'Leverage'.

Numa primeira abordagem, pode-se desde já concluir, a partir dos gráficos 1, 10, 11 e 12, relativos às medidas D_i (2.41), $DFITS_i$ (2.54), W_i (2.58) e T_i (2.59), respectivamente, apresentados no final desta subsecção, que no exemplo em estudo, elas não são contraditórias em termos ordinais. Apesar de estas medidas não fornecerem a mesma informação, a não contrariedade deve-se essencialmente ao facto de os valores das estimativas $\hat{\sigma}_{(i)}$, $i=1961, \dots, 1990$, não diferirem significativamente de $\hat{\sigma}^{11}$ (confronte-se os gráficos 3 e 4, no final

11 - É de realçar que, de acordo com Cook, R.D.(1986), D_i deve ser interpretado como uma medida da influência em β enquanto que $DFITS_i$ e T_i , devem ser interpretadas como medidas da influência em $(\beta, \hat{\sigma})$, como se referiu no final da secção 2.3. A interpretação da medida de Welsch, W_i , é dificultada pelo facto de a métrica utilizada depender da observação i .

desta subsecção, relativos aos resíduos studentizados t_i^2 e r_i^2 , respectivamente). Em consequência a identificação de eventuais observações influentes irá ser feito a partir da medida de Cook, não só pelo motivo apontado, mas também por ser uma das que mais se popularizou na literatura.

No que respeita a esta medida [veja-se o Quadro 3 seguinte ou o Gráfico 1, no final desta subsecção], constata-se que nenhuma das observações, tomadas individualmente, exerce uma influência considerável nos coeficientes estimados do modelo, pois os valores de D_i , $i=1961, \dots, 1990$, são significativamente inferiores à unidade. Apesar da influência diminuta, são de destacar os anos de 1973 e 1975 pelo valor assumido por p_u e t_i^2 , respectivamente [veja-se os gráficos 2 e 3]. Sendo a média dos elementos diagonais da matriz de projecção, P , dada por $k/n=0.1667$, a observação relativa a 1973 tem associada um valor de p_u ligeiramente acima do 'nível crítico' proposto por Hoaglin e Welsch (1978), $2(k/n)=0.333^{12}$. Quanto ao ano de 1975, pode-se observar pelo quadro 3 que a sua influência é em grande parte devida ao valor assumido por t_i que sob a hipótese nula é um 'outlier' não influente em β .

12 - Como tal, a linha da matriz X , referente ao ano de 1973, está relativamente afastada da média das linhas de X , isto é, do seu 'centro de gravidade'.

QUADRO 3

Maiores valores para D_i

ANO	D_i	p_a	t_i^2
1973	0.286	0.398	2.169
1989	0.261	0.257	3.780
1975	0.214	0.144	6.329
1986	0.140	0.137	4.406
1966	0.073	0.155	1.983
1983	0.061	0.142	1.836

Não havendo razões para se suspeitar da presença de observações influentes, procurou-se no entanto seguir o procedimento habitual de identificar quais os coeficientes mais afectados. Tendo presente os gráficos 5 a 9, no final desta subsecção, pode-se, para cada coeficiente, concluir o seguinte,

- β_0 o ano de 1973 é o mais influente, seguido do ano de 1966;
- β_1 o ano de 1975 é o mais influente seguidos dos anos de 1973 e 1986;
- β_2 o ano de 1973 é o mais influente seguidos dos anos de 1975 e 1966;
- β_3 o ano de 1973 é o mais influente seguidos dos anos de 1989 e 1966;
- β_4 o ano de 1973 é o mais influente seguido do

ano de 1975;

O ano de 1973, sendo o mais influente em β (apesar de esta não ser 'preocupante'), aparece também como o mais influente em cada um dos coeficientes do modelo.

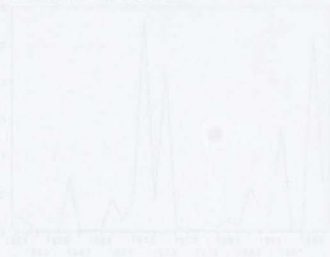


GRÁFICO 2

DIAGONAL DA MATRIZ DE PRELIMINAR β_1

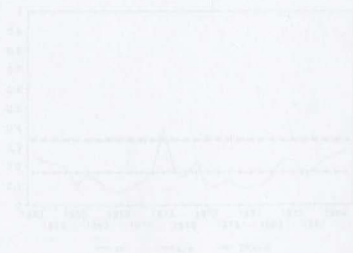


GRAFICO 1

MEDIDA DE COOK [D_i]

(Influência em β)

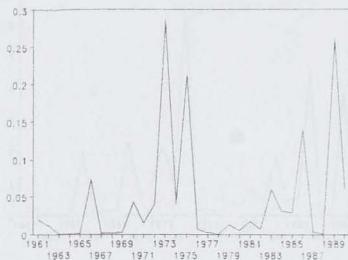


GRAFICO 2

DIAGONAL DA MATRIZ DE PROJECAO [p_{ii}]

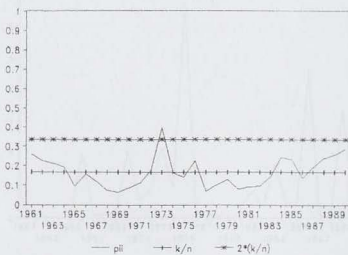


GRÁFICO 3

RESÍDUOS STUDENTIZADOS [t_i^2]



GRÁFICO 4

RESÍDUOS STUDENTIZADOS [r_i^2]

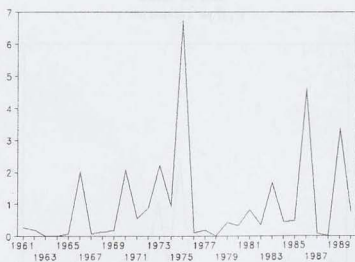


GRÁFICO 5

MEDIDA DE COOK
(Influência em β_0)

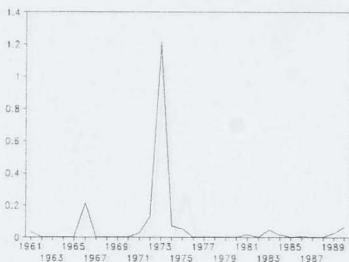


GRÁFICO 6

MEDIDA DE COOK
(Influência em β_1)

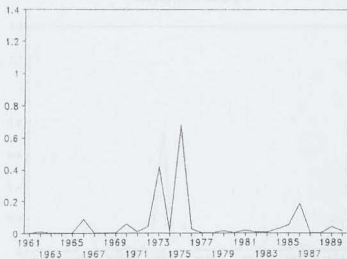


GRAFICO 7

MEDIDA DE COOK
(Influência em $\hat{\beta}_2$)

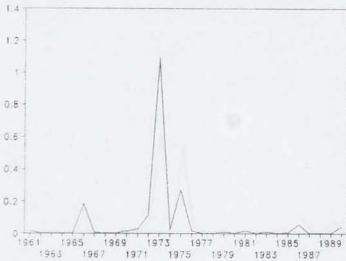


GRAFICO 8

MEDIDA DE COOK
(Influência em $\hat{\beta}_3$)

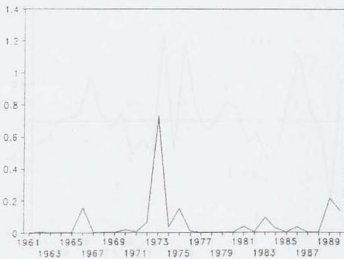


GRAFICO 9

MEDIDA DE COOK
(Influência em $\hat{\beta}_4$)

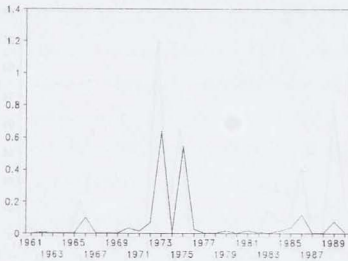


GRAFICO 10

MEDIDA DE WELSH-KUH [$DFITS_i$]
(Influência em $(\hat{\beta}, \hat{\sigma})$)

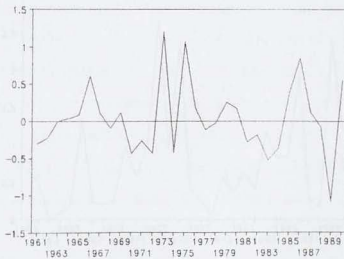


GRÁFICO 11

MEDIDA DE WELSH [W_i]

(Influência em $(\beta, \hat{\sigma})$)

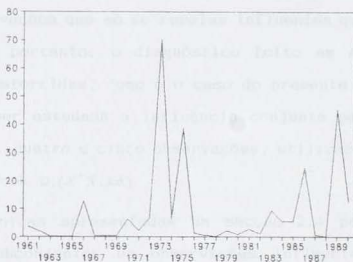
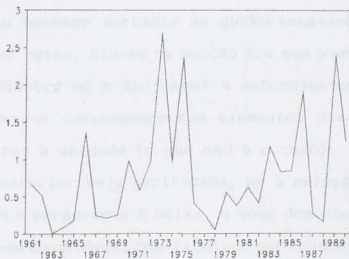


GRÁFICO 12

MEDIDA DE ATKINSON [T_i]

(Influência em $(\beta, \hat{\sigma})$)



B) Influência Conjunta.

Como se referiu no capítulo 2, secção 2.4, deste trabalho, existem observações que só se revelam influentes quando associadas com outras e portanto, o diagnóstico feito em A) pode levar a conclusões distorcidas, como é o caso do presente trabalho. Neste sentido vai ser estudada a influência conjunta para subconjuntos de duas, três, quatro e cinco observações, utilizando para o efeito a medida de Cook $D_i(X^T X, k\hat{\sigma})$.

As técnicas apresentadas na secção 2.4 para calcular os potenciais subconjuntos de observações influentes não são tão eficientes como seria de desejar. Por um lado, existe uma grande dificuldade em estabelecer um limite superior para estas medidas: se esse valor é reduzido corre-se o risco de deixar de fora alguns subconjuntos influentes; se é elevado, o número de subconjuntos potencialmente influentes aproxima-se do número de subconjuntos possíveis, não havendo portanto um ganho substancial em tempo de computação. Por outro, viu-se na secção 2.4 que o método apresentado por Cook e Weisberg só é aplicável a subconjuntos de observações em que a soma dos correspondentes elementos diagonais de P , P_i , sejam inferiores à unidade (o que não é o caso). Mas, mesmo que a desigualdade anterior seja verificada, se a relação entre o número de observações e parâmetros é baixa, a soma dos elementos diagonais de P_i aproxima-se da unidade, contribuindo para aumentar a dificuldade de detecção de potenciais subconjuntos influentes.

Como consequência destas dificuldades optou-se pelo cálculo da influência de todos os subconjuntos de observações, não deixando no entanto de salientar que o volume de cálculos e portanto o tempo de computação, aumenta consideravelmente com o número de elementos em cada subconjunto. No presente trabalho consideraram-se subconjuntos de 2,3,4 e 5 observações, resumindo-se no quadro seguinte o número de elementos e o tempo dispendido.

QUADRO 4

Tempo de Computação Dispendido¹³

Subconjuntos de m Obs.	Nº Elementos	Tempo (aproximado)
2	435	01'44''
3	4060	22'04''
4	27405	02h 19'
5	142506	13h 02'

Os resultados apresentados no quadro 4 falam por si. Para subconjuntos com mais do que 4 observações, o tempo dispendido tem um custo extremamente elevado, sendo aceitável para os de 2 ou 3 observações.

13 - Note-se que estes tempos dizem respeito ao tempo total (carregamento do programa, cálculos e registo dos resultados num ficheiro de output) e está obviamente condicionado pelo tipo de máquina e eficiência de código do programa.

Quanto à influência conjunta de duas observações, nenhum dos subconjuntos se apresenta influente em β , como se pode observar pelo quadro 5, ou pelo Gráfico 13, que a seguir se apresentam. É no entanto de realçar que os anos de 1988 e 1989, não evidenciados no diagnóstico individual, aparecem agora como os mais influentes em β .

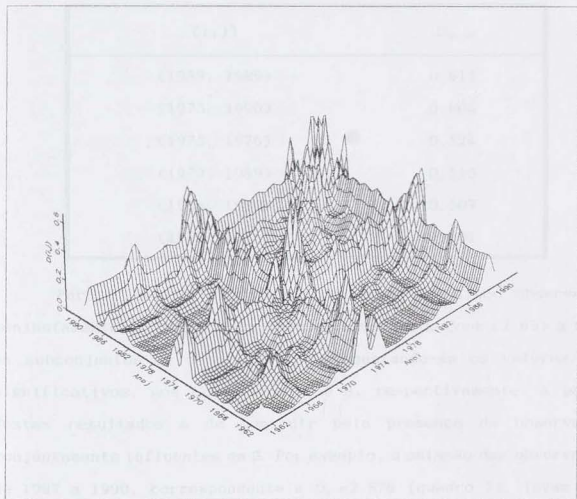
Utilizando o método de Gray e Ling para a determinação de potenciais observações influentes, são patentes as dificuldades inicialmente referidas. Uma dessas dificuldades, está no facto de $p_{x_{ij}}$ não ser uma transformação monótona de D , com $I=\{i,j\}$, $i,j=1961, \dots, 1990$ e $i \neq j$. Considerando, do mesmo modo que Gray e Ling, apenas os valores de $|p_{x_{ij}}| \geq 0.15$ (ver quadro 9) constata-se que não são considerados o 2º, 4º e 5º subconjuntos mais influentes. Por exemplo, o valor de $p_{x_{ij}}$ para $(i,j)=(1973,1989)$ é em valor absoluto bastante reduzido (0.07) e simultaneamente é identificado como o 4º subconjunto mais influente (ver Quadro 5). Se em contrapartida se tomassem os valores de $|p_{x_{ij}}| \geq 0.07$ haveria demasiados subconjuntos de duas observações a processar, prática que pode não ser vantajosa quando confrontada com o cálculo da influência de todos os conjuntos possíveis.

14 - Por conveniência de construção, o Gráfico 13 apresenta os resultados de $p_{x_{ij}}$ ($i,j=1961, \dots, 1990$, com $i \neq j$).

GRAFICO 13

Influência conjunta de 2 observações¹⁴

$$(D_{(i,j)}, i \neq j)$$



14 - Por conveniência de construção, o gráfico é simétrico em relação ao par (i,i) , $i=1961, \dots, 1990$, não incluído.

QUADRO 5

Influência conjunta de 2 observações

$$(D_{(i,j)} > 0.5)$$

(i,j)	$D_{(i,j)}$
(1988, 1989)	0.615
(1973, 1990)	0.604
(1975, 1976)	0.524
(1973, 1989)	0.513
(1986, 1990)	0.507
(1973, 1986)	0.501

Para a identificação de três, quatro e cinco observações conjuntamente influentes, aplicou-se a medida de Cook (2.65) a todos os subconjuntos de observações, apresentando-se os valores mais significativos, nos quadros 6, 7 e 8, respectivamente. A partir destes resultados é de concluir pela presença de observações conjuntamente influentes em β . Por exemplo, a omissão das observações de 1987 a 1990, correspondente a $D_i = 2.878$ (quadro 7), levam a um deslocamento da estimativa de β acima da região de confiança a 95% para β . É ainda interessante notar, a partir do quadro 8, a importância assumida pelas observações relativas aos anos de 1987 a 1990, juntamente com os anos de 1974, 1966, 1979, 1981, entre outros, que se revelaram pouco significativos no diagnóstico individual.

QUADRO 6

Influência conjunta de 3 observações

$$(D_{(i,j,k)} > 0.9)$$

(i,j,k)	$D_{(i,j,k)}$
(1973, 1986, 1990)	1.352
(1987, 1988, 1989)	1.127
(1966, 1975, 1976)	1.101
(1973, 1988, 1989)	1.021
(1985, 1986, 1990)	0.975
(1961, 1973, 1990)	0.925
(1973, 1983, 1984)	0.905
(1966, 1967, 1984, 1985)	1.6508
(1974, 1987, 1988, 1989)	1.8173
(1973, 1986, 1987, 1990)	1.3409
(1973, 1985, 1988, 1990)	1.3473
(1963, 1973, 1986, 1990)	1.3223
(1973, 1980, 1986, 1990)	1.3139
(1970, 1973, 1986, 1990)	1.3072

QUADRO 7

Influência conjunta de 4 observações

$$(D_{(i,j,k,l)} > 1.5)$$

(i, j, k, l)	$D_{(i, j, k, l)}$
(1987, 1988, 1989, 1990)	2.8783
(1973, 1987, 1988, 1989)	2.0515
(1973, 1985, 1986, 1990)	1.8778
(1966, 1967, 1975, 1976)	1.7769
(1973, 1979, 1986, 1990)	1.7201
(1961, 1973, 1986, 1990)	1.6918
(1964, 1973, 1986, 1990)	1.6699
(1986, 1987, 1988, 1990)	1.6598
(1974, 1987, 1988, 1989)	1.6175
(1973, 1986, 1987, 1990)	1.5489
(1973, 1986, 1988, 1990)	1.5473
(1963, 1973, 1986, 1990)	1.5224
(1973, 1980, 1986, 1990)	1.5139
(1970, 1973, 1986, 1990)	1.5022

QUADRO 8

Influência conjunta de 5 observações

$$(D_{(i,j,k,l,w)} \geq 3)$$

(i,j,k,l,w)	$D_{(i,j,k,l,w)}$
(1974, 1987, 1988, 1989, 1990)	5.867
(1966, 1987, 1988, 1989, 1990)	5.255
(1979, 1987, 1988, 1989, 1990)	4.190
(1981, 1987, 1988, 1989, 1990)	4.124
(1973, 1987, 1988, 1989, 1990)	3.510
(1976, 1987, 1988, 1989, 1990)	3.245
(1964, 1987, 1988, 1989, 1990)	3.215
(1982, 1987, 1988, 1989, 1990)	3.146
(1977, 1987, 1988, 1989, 1990)	3.100
(1983, 1987, 1988, 1989, 1990)	3.069
(1967, 1987, 1988, 1989, 1990)	3.063
(1969, 1987, 1988, 1989, 1990)	3.018
(1961, 1987, 1988, 1989, 1990)	3.001
(1978, 1987, 1988, 1989, 1990)	3.000

QUADRO 9 - MATRIZ DE PROJEÇÃO P_x
($|p_{xu}| \geq 0.15$)

	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
61	0.27	0.21	0.17												
62	0.21	0.23	0.21	0.18											
63	0.17	0.21	0.21	0.20											
64		0.18	0.20	0.20											
65															
66						0.22									0.20
67															
68															
69															
70										0.17					
71												0.16	0.15		
72												0.16	0.21	0.22	0.16
73												0.15	0.22	0.45	
74													0.16		0.20
75						0.20									0.36
76															0.21
77															
78															
79															
80															
81															
82															
83															
84															
85															
86										-0.15					
87															
88															
89															
90														0.14	

(Continua)

(Continuação)

	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
61															
62															
63															
64															
65															
66															
67															
68															
69															
70										-0.15					
71															
72															
73															
74															
75	0.21														
76	0.23														
77															
78															
79															
80															
81															
82															
83								0.20	0.21						
84								0.21	0.26	0.22					
85									0.22	0.25	0.20				
86										0.20	0.29				0.19
87												0.19	0.21	0.19	0.17
88												0.21	0.24	0.25	0.20
89												0.19	0.25	0.37	0.20
90											0.19	0.17	0.20	0.20	0.31

Avaliar as possíveis causas da influência não é tarefa fácil pois não existe uma solução comum para este problema. As causas podem ser diversas e devem ser analisadas caso a caso.

No que respeita ao problema em estudo passam a enumerar-se alguns dos possíveis factores:

- admitindo a validade do modelo com que se está a trabalhar, uma das razões da influência pode ter a ver com as séries económicas existentes. A passagem em 1977 do antigo para o novo sistema de contas nacionais trouxe problemas de compatibilização, por não haver um ano de reposição entre estes dois sistemas. Em consequência a compatibilização de algumas das séries foi baseada em processos 'expeditos', reduzindo a sua qualidade¹⁵;
- outro factor possível prende-se com o facto de a partir de 1985 se ter trabalhado com dados que não são definitivos, isto é, com dados, baseados em estimativas, extraídos dos Relatórios do Banco de Portugal;
- um terceiro e último factor pode resultar da combinação de uma deficiente qualidade dos dados com um modelo que pode não ser o mais adequado para explicar a procura de moeda em Portugal, por não ser capaz de assimilar toda a informação contida nas observações. Por exemplo, factores como o tipo de ajustamento à procura desejada,

15 - A este respeito veja-se Cartaxo e Rosa (1986).

4.2. CONCLUSÃO bem como o tipo de variáveis incluídas (e não incluídas, como é o caso da taxa de inflação e da taxa de juro externa), podem desempenhar aqui um papel importante.

Com base nestes possíveis factores, tendo presentes os valores da influência conjunta obtidos anteriormente, existem fortes suspeitas de que as observações posteriores a 1986 não tenham uma qualidade desejável. No entanto é importante frisar que esta conclusão carece de um estudo mais aprofundado.

Outros factores que se prendem com o tipo de modelo utilizado podem ter contribuído para explicar essa influência. A corroborar esta hipótese estão as diversas alterações na sociedade portuguesa em especial as posteriores aos anos de 1985/86,

- a integração na CEE,
- a progressiva liberalização dos mercados financeiros e
- o dinamismo do mercado de capitais.

Estas alterações, que não se esgotam nos três casos anteriores, podem ter contribuído para uma alteração do comportamento dos agentes económicos, face à procura de moeda, obrigando deste modo a um reespecificação mais cuidadosa do modelo (3.5).

4. CONSIDERAÇÕES FINAIS.

Uma grande parte dos trabalhos realizados em Econometria Aplicada assenta na utilização do estimador que não é robusto (mínimos quadrados). Em consequência, o valor das estimativas dos parâmetros (para além do modelo que lhes está subjacente), depende em grande parte da qualidade dos dados utilizados. Valores extremos no espaço das variáveis do modelo, podem levar a uma alteração substancial dessas estimativas. Como tal, é desejável que na fase de avaliação de qualquer modelo econométrico se disponha de alguma medida que permita avaliar sobre a qualidade (estabilidade) dos seus parâmetros, em função dos dados disponíveis.

Neste contexto, o estudo da influência das observações surge como um método de diagnóstico extremamente importante em qualquer trabalho de econometria aplicada, apesar de ainda hoje ser pouco utilizado.

Enquanto o estudo da influência individual, nas suas múltiplas abordagens, está convenientemente tratado na literatura, o mesmo não se pode afirmar no que diz respeito à influência conjunta, como se viu no capítulo 3. Uma primeira dificuldade prende-se com a dimensão do problema, tornando extremamente custosa a análise de todos os casos possíveis. Uma segunda tem a ver com o facto de as técnicas apresentadas por Cook e Weisberg (1980) e Gray e Ling

(1984), com o propósito de obviar as dificuldades anteriores, não se revelaram suficientemente eficazes, em especial quando se trabalha com uma amostra pouco reduzida.

Por este facto, o estudo da influência conjunta, em especial a partir de meados dos anos 80, tem sido feito com o recurso à utilização de técnicas robustas, mais eficazes e menos dispendiosas, por autores como Atkinson (1986), Kauss e Bradu (1984) e Rousseeuw e Leroy (1987).

Calculo da influencia conjunta no modelo para o primeiro modelo
1961-1970

Leitura do ficheiro de dados

DATA DO DIA

DATA DO DIA

DATA DO DIA

DATA DO DIA

Informação de variáveis

DATA DO DIA

DATA DO DIA

DATA DO DIA

DATA DO DIA

5. ANEXOS.

5.1 PROGRAMA 'TSP' PARA O CALCULO DA INFLUÊNCIA CONJUNTA.

Apresenta-se de seguida o programa, em TSP, que serviu de base para o cálculo da influência conjunta de 2, 3, 4 e 5 observações.

```
?  
?  
?  
?  
?  
? Cálculo da Influência conjunta no modelo para a procura Moeda  
1961-1990
```

```
?  
?  
?  
? Leitura do ficheiro de dados.  
?
```

```
OPTIONS CRT;  
FREQ N;  
SMPL 1960 1990;  
LOAD (File='MOEDA.DAT') y jdp p m1;
```

```
?  
?  
?  
? Transformação de variáveis.  
?
```

```
GENR Lp=log(p);  
GENR Lm1=log(m1);  
GENR Ljdp=log(jdp);  
GENR Ly=log(y);
```

?
 ? Estimação do modelo
 ?

SMPL 1961 1990;
 TREND (start=1961,step=1) t;
 LIST vdi lml c Lp Ly Ljdp lml(-1); ? Lista com as variáveis do modelo

OLSQ vdi; ? Estimação pelo MOMQ.

SET sig2=@s2; ? Variância.
 MFORM Beta=@coef; ? Vector dos Coeficientes estimados.
 MMAKE X c Lp Ly Ljdp lml(-1); ? Matriz X dos regressores.
 MAT M=X'X;

SUPRES @logl @ses @coef @smpl;

inf2 vdi Beta sig2; ?
 inf3 vdi Beta sig2; ? Procedimentos para
 inf4 vdi Beta sig2; ? cálculo da influência conjunta.
 inf5 vdi Beta sig2; ?

?
 ? PROCEDIMENTOS
 ?

?
 ? Influência de duas observações.
 ?

PROC inf2 vars B s2;
 SET cont=0;
 SET a=1961;
 DO i=1961 to 1989 by 1;
 SET a=a+1;
 DO j=a to 1990 by 1;
 SET cont=cont+1;
 SMPLIF t^=i & t^=j;
 OLSQ (silent) vars;
 MAT Cook = (@coef-B)'M*(@coef-B)/(@coef*s2);
 IF Cook > 0.5; THEN; DO;
 PRINT i j Cook;
 ENDDO;
 SMPL 1961 1990;
 ENDDO;
 ENDDO;
 PRINT cont;
 ENDPROC;

2
?
?-----
Influência de três observações.

```
PROC inf3 vars B s2;  
  SET cont=0;  
  SET a1=1961;  
  SET a2=1962;  
  DO i=1961 TO 1988 BY 1;  
    SET a1=a1+1;  
    DO j=a1 TO 1989 BY 1;  
      SET a2=a2+1;  
      DO k=a2 TO 1990 BY 1;  
        SET cont=cont+1;  
        SMPLIF t^=i & t^=j & t^=k;  
        OLSQ (silent) vars;  
        MAT Cook = (@coef-B)'M*(@coef-B)/(@ncoef*s2);  
        IF Cook > 0.8; THEN; DO;  
          PRINT i j k Cook;  
        ENDDO;  
        SMPL 1961 1990;  
      ENDDO;  
    ENDDO;  
  SET a2=a1+1;  
ENDDO;  
PRINT cont;  
ENDPROC;
```

 ? Influência de quatro observações.
 ?-----

```

PROC inf4 vars B s2;
  SET cont=0;
  SET b1=1961;
  SET b2=1962;
  SET b3=1963;
  DO i=1961 TO 1987 BY 1;
    SET b1=b1+1;
    DO j=b1 TO 1988 BY 1;
      SET b2=b2+1;
      DO k=b2 TO 1989 BY 1;
        SET b3=b3+1;
        DO l=b3 TO 1990 BY 1;
          SET cont=cont+1;
          SMPLIF t^=i & t^=j & t^=k & t^=l;
          OLSQ (silent) vars;
          MAT Cook = (@coef-B)'M*(@coef-B)/(@ncoef*s2);
          IF Cook > 0.99; THEN; DO;
            PRINT i j k l Cook;
          ENDDO;
          SMPL 1961 1990;
        ENDDO;
      ENDDO;
    SET b3=b2+1;
  ENDDO;
  SET b2=b1+1;
  SET b3=b2+1;
ENDDO;
PRINT cont;
ENDPROC;

```

Influência de cinco observações.

```

PROC inf5 vars b s2;
  SET cont=0;
  SET b1=1961;
  SET b2=1962;
  SET b3=1963;
  SET b4=1964;
  DO i=1961 TO 1986 BY 1;
    SET b1=b1+1;
    DO j=b1 TO 1987 BY 1;
      SET b2=b2+1;
      DO k=b2 TO 1988 BY 1;
        SET b3=b3+1;
        DO l=b3 TO 1989 BY 1;
          SET b4=b4+1;
          DO w=b4 TO 1990 BY 1;
            SET cont=cont+1;
            SMPLIF t^=i & t^=j & t^=k & t^=l & t^=w;
            OLSQ (silent) vars;
            MAT Cook = (@coef-B)'M*(@coef-B)/(@ncoef*s2);
            IF Cook > 1.5; THEN; DO;
              PRINT i j k l w Cook;
            ENDDO;
            SMPL 1961 1990;
          ENDDO;
        ENDDO;
      SET b4=b3+1;
    ENDDO;
  SET b3=b2+1;
  SET b4=b3+1;
ENDDO;
SET b2=b1+1;
SET b3=b2+1;
SET b4=b3+1;
ENDDO;
PRINT cont;
ENDPROC;

```

5.2 DADOS ESTATÍSTICOS.

ANO	Deflactor Consumo Privado	PIB pm P.77	Taxa de Juro dos Depósitos a Prazo 6 Meses/ 1 Ano	MOEDA M1
	(p)	(y)	(jdp)	(ml)
1960	0.28304	246.037	1.250	40.737
1961	0.28461	258.831	1.250	41.276
1962	0.29031	275.914	1.250	44.673
1963	0.29356	292.192	1.250	50.214
1964	0.29576	312.938	1.250	56.823
1965	0.31001	336.096	2.375	62.198
1966	0.32692	348.867	3.500	70.148
1967	0.33192	377.125	3.625	75.274
1968	0.34624	404.656	4.000	80.725
1969	0.36327	418.009	4.000	91.320
1970	0.37488	453.958	5.125	92.582
1971	0.40101	483.920	5.750	104.900
1972	0.42635	522.633	5.750	120.928
1973	0.46417	581.168	5.750	165.526
1974	0.57343	587.561	6.250	182.254
1975	0.66522	562.296	8.000	227.098
1976	0.78577	597.158	9.500	246.319
1977	1.00000	625.835	11.625	274.807
1978	1.21200	647.033	17.000	314.878
1979	1.50400	689.737	19.000	396.569
1980	1.79400	718.016	19.000	484.536
1981	2.18700	721.606	19.250	524.597
1982	2.67876	744.697	20.825	610.601
1983	3.36153	742.463	25.700	666.375
1984	4.34641	730.584	28.000	772.619
1985	5.17232	754.693	25.500	980.701
1986	5.87058	787.145	17.400	1334.402
1987	6.47525	827.289	14.500	1527.179
1988	7.10335	860.381	13.200	1722.082
1989	8.02679	907.702	13.000	1827.821
1990	9.10238	947.641	13.547	2351.671

Unidades e Fontes:

- p: Índice de base fixa (1977=1). Cartaxo e Rosa (1986) e Relatórios do Banco de Portugal.
- y: Preços constantes de 1977, 10^6 Contos, Cartaxo e Rosa (1986) e Relatórios do Banco de Portugal.
- jdp: Percentagens, Banco de Portugal (Dados não Publicados).
- ml: Preços Correntes, 10^6 Contos, Cartaxo e Santos (1984) e Relatórios do Banco de Portugal.

6. BIBLIOGRAFIA

- Andrews, D. F. e Pregibon, D. (1978), "Finding the Outliers that Matter", *Journal of the Royal Statistical Society B*, 40, 85-93.
- Atkinson, A. C. (1981), "Two Graphical Displays for Outlying and Influential Observations in Regression", *Biometrika*, 68, 13-20.
- Atkinson, A. C. (1986), "Masking Unmasked", *Technometrics*, 73, 533-541.
- Behnken, D. W. e Draper, N. R. (1972), "Residuals and their Variance Patterns", *Technometrics*, 14, 102-111.
- Belsley, David A.; Kuh, Edwing e Welsch, Roy E. (1980), *Regression Diagnostics*, New-York: John Wiley & Sons.
- Cartaxo, R. J. e Rosa, N. E. S. (1986), Séries Longas para as Contas Nacionais Portuguesas 1958-1985. Banco de Portugal, Gabinete de Estudos, *Documento de Trabalho N.º 15*.
- Cartaxo, R. J. e Santos, E. A. (1984), Estimativas Anuais da Riqueza Financeira das Famílias para o período 1958-1981, Banco de Portugal, Gabinete de Estudos, *Documento de Trabalho N.º 8*.
- Chatterjee, Samprit e Hadi, Ali S. (1986), "Influential Observations, High Leverage Points, and Outliers in Linear Regression", *Statistical Science*, 1, 379-416.
- Chatterjee, Samprit e Hadi, Ali S. (1988), *Sensitivity Analysis in Linear Regression*, John Wiley & Sons.
- Chow, G. C. (1966), "On the Long-Run and Short-Run Demand For Money", *Journal of Political Economy*, 74, 111-131.

- Cook, R. Dennis (1977), "Detection of Influential Observation in Linear Regression", *Technometrics*, 19, 15-18.
- Cook, R. Dennis (1979), "Influential Observations in Linear Regression", *Journal of the American Statistical Association*, 74, 169-174.
- Cook, R. Dennis (1986), "Comment on Influential Observations, High Leverage Points, and Outliers in Linear Regression", by Chatterjee e Hadi (1986), *Statistical Science*, 1, 393-397.
- Cook, R. Dennis e Weisberg, Sanford (1980), "Characterizations of an Empirical Influence Function for Detecting Influential Cases in Regression", *Technometrics*, 22, 495-508.
- Cook, R. Dennis e Weisberg, Sanford (1980), *Residuals and Influence in Regression*, Chapman and Hall.
- Cuthbertson, K. e Taylor, M. P. (1987), "The Demand for Money: A Dynamic Rational Expectations Model", *The Economic Journal*, Supplement, 97, 65-76.
- Davidson, J. E. H.; Hendry, D. F.; Srba, F.; Yeo, S. (1978), "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers Expenditure and Income in the United Kingdom", *The Economic Journal*, 88, 661-692.
- Davies, H. A. (1981), *Order Statistics*, 2nd Edn, New York: Wiley.
- Davies, R. B. e Hutton, B. (1975), "The Effects of Errors in the Independent Variables in Linear Regression", *Biometrika*, 62, 383-391.

- Ellenberg, J. H. (1973), "The Joint Distribution of the Standardized Least Squares Residuals from a General Linear Regression", *Journal of the American Statistical Association*, 68, 941-943.
- Engle, Robert F. (1982), "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation", *Econometrica*, 50, 987-1007.
- Fernholz, Luisa Turrin (1983), *Von Mises Calculus for Statistical Functionals*, Lectures Notes in Statistics, 19, Springer-Verlag.
- Gentleman, J. F. e Wilk, M. B. (1975), "Detecting Outliers in Two-Ways Tables I: Statistics an Behavior of Residuals", *Technometrics*, 17, 1-14.
- Goodhart, C. A. E. (1989), *Money Information an Uncertainty*, 2nd ed. Macmillan.
- Gray, J. B. e Ling, R. F. (1984), "K-Clustering as a Detection Tool for Influential Subsets in Regression", *Technometrics*, 26, 305-330.
- Hampel, F. R. (1968), Contributions to the Theory of Robust Estimation. Unpublished *Ph.D. Dissertation*, University of California, Berkeley, Sumariado em Hampel (1974), op.cit.
- Hampel, F. R. (1974), "The Influence Curve and its Role in Robust Estimation", *Journal of the American Statistical Association*, 62, 1179-1186.
- Harvey, Andrew (1982), *The Econometric Analysis of Time Series*, Deddington, Philip Allan.
- Hawkins, D. M. (1980), *Identification of Outliers*. London: Chapman and Hall.

- Hawkins, D. M.; Bradu, D. e Kass, G. (1984), "Location of Several Outliers in Multiple-Regression Data Using Elemental Sets", *Technometrics*, 26, 197-208.
- Henderson, H. V. e Searle, S. R. (1981), "On Deriving the Inverse of a Sum of Matrices", *SIAM Review*, 23, 53-60.
- Hendry, D. F. (1980), "Econometrics - Alchemy or Science ?", *Economica*, 47, 387-406.
- Hinkley, D. V. (1977), "Jackknifing in Unbalanced Situations", *Technometrics*, 19, 285-292.
- Hoaglin, D. C. e Welsch, R. E. (1978), "The Hat Matrix in Regression and ANOVA", *The American Statistician*, 32, 17-22.
- Huber, P. (1973), "Robust Regression: Asymptotics Conjectures and Monte Carlo", *Annals of Statistics*, 1, 799-821.
- Huber, P. (1981), *Robust Statistics*. John Wiley & Sons.
- Jaeckel, L. A. (1972), "The Infinitesimal Jackknife", Unpublished Bell Telephone Laboratories Report, Murray Hill, N.J., Sumariado em Miller (1974), op.cit.
- Johnson, W. e Geisser, S. (1979), "Assessing the predictive influence of observations", University of Minnesota, *School of Statistics Technical Report N° 355*.
- Johnson, W. e Geisser, S. (1980), "A predictive View of the Detection and Characterization of Influential Observations in Regression Analysis", University of Minnesota, *School of Statistics Technical Report N° 365*.
- Johnston, J. (1985), *Econometric Methods*, 3ª ed., MacGraw-Hill: International Student Edition.

- Kiviet, J. F. (1985), "Model Selection Test Procedures in a Single Linear Equation of Dynamic Simultaneous Equation System and Their Defects in Small Samples", *Journal of Econometrics*, 28, 327-362.
- Kiviet, J. F. (1986), "On the Rigour of Some Misspecifications Tests for Modelling Dynamic Relationships", *The Review of Economic Studies*, 53, 241-261.
- Kramer, W. e Sonnberger, H. (1986), *The Linear Regression Model Under Test*, Physica-Verlag, Heidelberg, Germany.
- Ling, R. F. (1972), "On the Theory and Construction of K-Clusters", *Computer Journal*, 15, 326-332.
- Luís, Gregório e Ribeiro, C. Silva (1985), *Algebra Linear*, MacGraw-Hill: Portugal.
- Margolin, Barry H. (1977), "The Distribution of Internally Studentized Statistics via Laplace Transform Inversion", *Biometrika*, 64, 573-582.
- Miller, R. (1974), "The Jackknife - A Review", *Biometrika*, 61, 1-16.
- Montgomery, D. C. e Peck, A. E. (1982), *Introduction to Linear Regression Analysis*, John Wiley & Sons.
- Murteira, Bento (1990), *Probabilidades e Estatística*, vol. I e II, 2ª ed., MacGraw-Hill: Portugal.
- Pollock, D. S. G. (1979), *The Algebra of Econometrics*. John Wiley & Sons.
- Pregibon, D. (1979), *Data Analytic Methods for Generalized Linear Models*, Unpublished *Ph.D. dissertation*, University of Toronto, Sumariado em Pregibon (1981), op.cit.

- Pregibon, D. (1981), "Logistic Regression Diagnostics", *Annals of Statistics*, 9, 705-724.
- Rao, C. R. (1973), *Linear Statistical Inference and Its Applications*, 2nd ed., John Wiley & Sons: New York.
- Rousseeuw, P. J. (1984), "Least Median of Squares Regression", *Journal of the American Statistical Association*, 79, 871-880.
- Rousseeuw, P. J. e Leroy, A. M. (1987), *Robust Regression & Outlier Detection*, John Wiley & Sons: New York.
- Sen, Ashish e Srivastava, Muni (1990), *Regression Analysis*, Springer-Verlag
- Serfling, Robert J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons.
- Velleman, P. e Welsch, R. (1981), "Efficient Computing of Regression Diagnostics" *American Statistician*, 35, 234-242.
- Von Mises, R. (1947), "On the Asymptotic Distributions of Differentiable Statistical Functions", *Annals of Mathematical Statistics*, 18, 309-348.
- Weisberg, Sanford (1980), *Applied Linear Regression*, John Wiley & Sons.
- Welsch, R. E. (1982), "Influence Functions and Regression Diagnostics" in *Modern Data Analysis*, R. L. Launer and A. F. Siegel, eds., New york: Academic Press.
- Welsch, R. E. e Kuh, Edwin (1977), "Linear Regression Diagnostics", Sloan School of Management, *Working Paper*, 923-977, M.I.T., Cambridge, Massachusetts.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica*, 48, 817-838.